

Designing and conducting health system research projects, *volume 2*, Data analyses and report writing

Corlien M. Varkevisser
Indra Pathmanathan
Ann Brownlee

KIT Publishers
International Development
Research Centre

Designing and Conducting Health Systems Research Projects

Volume II: Data analysis and report writing

This page intentionally left blank

Designing and Conducting Health Systems Research Projects

Volume II: Data analysis and report writing

Corlien M. Varkevisser

Indra Pathmanathan

Ann Brownlee

**KIT Publishers, Amsterdam
International Development Research Centre**

**in association with
WHO Regional Office for Africa**

Designing and conducting health systems research projects

Volume II: Data analysis and report writing

Jointly published by KIT Publishers and the International Development Research Centre (IDRC), in association with the Africa Regional Office (AFRO) of the World Health Organization.

KIT Publishers
Mauritskade 63, 1090 HA Amsterdam, the Netherlands
publishers@kit.nl / www.kit.nl
ISBN 90 6832 148 X

International Development Research Centre
PO Box 8500, Ottawa, ON, Canada K1G 3H9
info@idrc.ca / www.idrc.ca
ISBN 1-55250-069-1 (Volume 1) / 1-55250-070-5 (Volume 2)

World Health Organization - Regional Office for Africa
Cite du Djoue, P.O.Box 06 Brazzaville, Congo
www.whoafr.org

All rights reserved. No part of this publication may be reproduced stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior permission of KIT Publishers and the International Development Research Centre.

Cover design: Mulder en van Meurs, Amsterdam
Production: Meester & de Jonge, Lochem
NUR 600

TABLE OF CONTENTS

Foreword	vii
Preface and acknowledgements	ix
Introduction to Part II: Data Analysis and Report Writing	xi
Example of a Course Schedule for the Workshop on Data Analysis and Report Writing	xiii
Module 21: Orientation to the workshop on data analysis and report writing	1
Module 22: Description of variables	15
Module 23: Analysis of qualitative data	33
Module 24: Cross-tabulation of quantitative data	51
Module 25: Measures of association based on risk	65
Module 26: Dealing with confounding variables	81
Module 27: Preparation for statistical analysis: Measures of dispersion, normal distribution and sample variation	95
Module 28: Choosing a significance test	109
Module 29: Determining differences between groups: Part I, Analysis of unpaired observations	125
Module 30: Determining differences between groups: Part II, Analysis of paired observations	143
Module 31: Measuring associations between variables: Regression and correlation	153
Module 32: Writing a research report	167
Module 33: Dissemination, communication and utilisation of research findings	183
About the authors	195

This page intentionally left blank

FOREWORD

Health Systems Research (HSR) has proved to be a useful tool for health decision makers at all levels over the past 20 years, providing them with the necessary data for informed decision making.

The Joint HSR Project for the Southern African Region based in the WHO Office in Harare and supported by WHO Geneva, the Royal Tropical Institute (KIT) in Amsterdam and the Dutch Technical Development Co-operation (DGIS), has played a crucial role in the promotion of HSR in the African region since 1987. HSR was enthusiastically embraced by many Ministries of Health and universities. In 1996, the Regional WHO Office for Sub-Saharan Africa (AFRO) assumed full responsibility for implementing HSR. Following the recommendation of Health Ministers of the Region, WHO/AFRO in 1998 included HSR as a regular programme for all its 46 member states.

The present HSR training modules, developed by an interdisciplinary, international team of practical researchers, have been highly instrumental in raising the interest for HSR. Originally designed for health managers at different levels as a tool to develop problem solving research in the Southern African Region, the modules also proved useful in Malaysia and were further elaborated by staff of the School of Public Health The 1991 combined version, published by International Development Research Centre, Canada and WHO, Geneva,* was translated in French, Spanish and Portuguese, and sections of it appeared in Arabic, Vietnamese and Chinese. In different parts of the world the modules facilitated the development and implementation of hundreds of research protocols by health staff and researchers. The HSR modules are used in the Community Health and Social Science Departments of many African, Asian and Latin American universities to train students and prepare them for their fieldwork. They are also used by Masters of Public Health courses in Europe and the USA and by international research programmes interested in applied research.

This unanticipated application of the modules in academic as well as health management circles led to the rapid exhaustion of the 1991 edition and the several subsequent reprints. Various groups of users made many useful suggestions for changes and improvements. The HSR Unit in AFRO, with agreement from IDRC, therefore decided to organise a revision of the HSR modules. An interdisciplinary group of Southern African researchers reviewed and made revisions in two workshops in 1998 and 1999. Two of the three original editors finalised the present version. IDRC took on the final responsibility for the publication, which was financed by AFRO and IDRC and published by KIT.

It is hoped that this revised version of the modules will fulfil the same need as preceding ones have done. Certainly many new and persisting health problems urgently require operational research. How to support necessary health reforms and at the same time ensure equity in access to health care for high-risk groups remains a major challenge. HSR is one of the tools we have to obtain deeper insight in these challenges and optimally focus our resources.

Dr. Rufaro R. Chatora, Director of the Division of Health Systems and Services Development (DSD), WHO/AFRO, Harare

Dr. Christina Zarowsky, Senior Health Specialist, IDRC, Ottawa

Ms. Catherine Hodgkin, Head Health Department, KIT, Royal Tropical Institute, Amsterdam

* Corlien M. Varkevisser, Indra Pathmanathan and Ann Brownlee (1991) *Designing and conducting Health Systems Research projects. Part I: Proposal development and fieldwork; Part II: Data analysis and report writing*. Ottawa: Health Sciences Division of the International Development Research Centre (IDRC) and Geneva: Programme on Health Systems Research and Development of the World Health Organisation.

This page intentionally left blank

PREFACE AND ACKNOWLEDGEMENTS

The present volume *'Designing and conducting Health Systems Research'* is a thorough revision of Volume 2 of the *Health Systems Training Series* which the International Development Research Centre (IDRC) in Canada and WHO HQ in Geneva published in 1991 and reprinted several times under the same name. It became necessary to revise the modules, because over the years inevitable shortcomings and gaps were detected which needed to be addressed. been added.

Health managers, for example, stressed that implementation of the research findings and recommendations were somewhat underexposed in the modules. This point is now taken care of in Module 1 by adding a fourth, implementation phase to the Health Systems Research (HSR) training cycle which initially consisted of three phases: HSR proposal development (15 days), fieldwork (roughly 6 months) and data analysis and report writing (2 weeks). The implementation of research findings and recommendations is further elaborated in Module 33. Furthermore, health managers pleaded, understandably, for shorter courses. This wish has been taken care of by stressing more explicitly in Modules 1 and 3, as well as in the Course Guidelines (annexed to Part 1 of this volume) that the proposal development phase can be shortened by having research teams select their research topic in the field before the onset of the course, preferably under guidance of a facilitator. In addition, the WHO/AFRO HSR Programme based in Harare, is at present developing modules for participatory rapid action in health research at health centre and district levels which can be carried out and integrated in the day to day activities of staff and community members.

Research staff from Community Health, Social Science and other university departments/ research institutes in Sub-Saharan Africa or other parts of the world who are using the modules had other wishes. They advocated that, in addition to the already well-emphasized problem-solving, analytical research approaches, more weight should be given to descriptive research. A descriptive diagram has therefore been added to the problem analysis diagram in Module 4. In all subsequent research steps, if relevant, the distinction between analytic and descriptive studies has been elaborated. Qualitative research methods have also been given more weight and they were more thoroughly integrated with quantitative methods in the research methodology (Modules 8-14). This applies, for example, to Modules 10 (Data collection techniques) and 11 (Sampling techniques). Furthermore, two new modules have been added to Part 2 (Data analysis and report writing) of the volume: one on Measures of association based on risk (Module 25), which used parts of Module 30 in the 1991 version, and one on the difficult issue of Confounding variables (Module 26). This need for extension was also reflected in the most recent evaluation of HSR training (1997).*

Facilitators, finally, desired more elaborate examples of crucial research and data-analysis techniques. Therefore, Module 10B (Development of research instruments) has been elaborated with a section on interview techniques with interview exercises, and Module 10C (FGDs) now contains an example of a transcribed focus group discussion with codes in the margin. To Module 13 (Plan for data analysis), an example of a full-fledged questionnaire and of a master sheet have been added, and Module 23 (Analysis of qualitative data) now provides an example of a filled-in compilation sheet. Module 5 (Literature review), has been extended with an example of a literature review.

Apart from these additions, in all modules parts that had proven to be unclear or incomplete were rewritten, and many examples and references were replaced by more recent ones or elaborated.

The present revision was initiated in a workshop held from 2-11 November 1998 in Arusha by a group of interdisciplinary researchers and managers convened by the manager of the WHO/AFRO HSR Programme (since 1992 Gabriel Mwaluko). All participants had thorough experience with the modules and with HSR: Sambe Duale, Lawrence Gakuri, Pilate Khulumani, Steve Kinoti, Gabriel Mwaluko, Jude Padayachi, Brian Pazvakavambwa, Corlien Varkevisser and Godfrey Woelk. In August 1999 a group of three people (Alasford Ngwengwe, Corlien Varkevisser and Godfrey

* Corlien M. Varkevisser, Indra Pathmanathan and Ann Brownlee (1991) *Designing and conducting Health Systems Research projects. Part I: Proposal development and fieldwork; Part II: Data analysis and report writing*. Ottawa: Health Sciences Division of the International Development Research Centre (IDRC) and Geneva: Programme on Health Systems Research and Development of the World Health Organisation.

Woelk) made further revisions and synchronised the different texts in the WHO/AFRO/HSR office in Harare, supported by staff of the HSR office (since 1999 headed by Isabel R. Aleta, with Makhmokha Mohale and Eric Naterop as APOs). Corlien Varkevisser and Ann Brownlee finalised and edited the modules, with the blessing of Indra Pathmanathan who this time could not participate. Deborah Karugonjo (Harare) and Merel Gallée (Amsterdam) provided highly valued assistance in the production of successive computerised versions. Funds for revising and publishing the HSR modules were made available by DGIS (Dutch Development Co-operation); SARA/AED, Washington; GTZ, Germany through the GTZ MCH/FP network for Health Systems Research in Southern Africa; WHO/AFRO and by WHO HQ, Geneva. IDRC, Canada assists in subsidised distribution of the modules.

A highly varied collection of people assisted in the production of earlier versions of the HSR modules. The cradle of the modules stood in Western Africa, where in the early eighties the Project for Strengthening Health Delivery Systems (SHDS), based in Boston University, USA, at the request of AFRO developed training materials in research protocol development. SHDS followed the step-by-step approach which till today is a major key to the success of HSR courses. Modules 1-17 in this volume are heavily adapted or new versions of the original SHDS modules.* The first adaptation took place in 1988, with 12 researchers from countries that participated in the Joint HSR Project (Omondi (Kenya), Sebatane and Makatjane (Lesotho), Chimimba and Msukwa (Malawi), Kitua and Savy (Seychelles), Tembo (Zambia) Munochiveyi, Taylor and Woelk (Zimbabwe) and Joint Project staff which also finalised the version (Corlien Varkevisser and Martien Borgdorff). These 'green modules'* found their way to Malaysia, where Indra Pathmanathan further developed them, with assistance from Maimunah Abdul Hamid, K. Mariappan and C. Sivagnanasundram (Sri Lanka), in the course of numerous protocol development workshops. The same occurred in Southern and Eastern Africa. At the initiative of Yvo Nuyens, who fathered the Joint HSR Project in WHO Geneva, and supported by IDRC (Annette Stark), the five volumes of the *Health Systems Research Training Series* emerged, of which *Designing and Conducting Health Systems Research Projects* formed Volume 2. These 'pink modules', published in 1991 in Ottawa by IDRC and WHO, form a thorough merge of the ever developing Southern African and Malaysian versions. They were integrated in Harare (Corlien Varkevisser and Leon Bijlmakers), in consultation with Indra Pathmanathan, and with thorough editing support from Ann Brownlee, one of the authors of the original SHDS modules. The present HSR modules are therefore a truly global production. It is even impossible to mention every contributor, because many HSR course facilitators and participants through their questions and critical remarks inspired further changes.

With such a colourful and interactive origin it seems highly unlikely that the present reprint will be the last one. Whenever the modules are used, they will be adapted. We hope, however, that in their present form they will last for some years and will be of use to health staff as well as university students.

Dr. Corlien M. Varkevisser, Royal Tropical Institute/University of Amsterdam

Dr. Ann Brownlee, University of California, San Diego

June 2003

* Regional Assessment of Health Systems Research Training in Eastern and Southern Africa. HSR Project and SARA/AED, Harare/Washington. SS Ndeki, 1997.

* Ann Brownlee, Thomas Nchinda and Yolanda Mousseau-Gershman (1983) *Health Services Research Course: How to develop proposals and design research to solve priority problems*. Boston: Boston University Health Policy Institute.

* Joint World Health Organisation/Royal Tropical Institute/Dutch Technical Development Co-operation Project on Health Systems Research for the Southern African Region (1988) *Health Systems Research Training Course: How to develop research proposals to solve priority health problems*. Geneva: World Health Organisation. WHO/SHS/HSR/88.3.

INTRODUCTION TO PART II: Data Analysis and Report Writing

This publication is meant to be used in combination with Part I: *Proposal Development and Fieldwork*. Part I consists of 20 training modules which, step-by-step, support course participants in the development of a research proposal and provide useful guidelines for its implementation.

The present Part II, *Data Analysis and Report Writing*, consists of 13 modules. These training modules on data analysis, report writing, and planning for implementation of recommendations, to a still larger extent than those on the development of a research proposal, can be used in a flexible way, depending on:

- the educational level and research experience of the course participants;
- the type of study conducted and type(s) of data collection techniques used; and
- the state in which the data are at the onset of the data analysis and report writing workshop.

If participants have some previous training in research methodology and statistics, and research experience, presentations of modules may be short. In this case the purpose of presenting is mainly to refresh the participants' memories and to guide them towards correct application of appropriate analysis procedures and tests. Some modules may then be combined or shortened.

If participants have neither training nor experience in research, the presentation of the materials in the modules may have to be restricted to the bare essentials required to handle the data that has been collected. Under these circumstances presentations may take longer and should include ample opportunity for asking questions and for classroom exercises.

'Bare essentials' that could be considered include:

- **Module 21** (Orientation to the workshop).
- **All of Modules 22 and 24** (Description of variables and cross tabulation).
- **Module 23** (Analysis of qualitative data, especially **parts I to IV**).
- **Module 25** (Measures of association based on risk: incidence, risk, relative risk and odds ratio). Concentrate on unpaired observations.
- **Briefly: Module 26** (Dealing with confounding variables). What confounding is, and how to deal with it should become clear through some examples.
- **Module 27** (Preparation for statistical analysis: measures of dispersion, normal distribution and sample variation).
- **Module 28** (Choosing a significance test). Concentrate globally on **parts I, II and III**, explaining the rationale for significance tests and how they work, but deal only briefly with part IV, the actual choosing of a significance test, if the groups are not likely to use more than the chi-square and/or the t-test.
- **Module 29** (Determining differences between groups: analysis of unpaired observations). Either the t-test or the chi-square test or both.
- **All of Modules 32** (Report writing) **and 33** (Promoting the dissemination, communication and utilisation of research findings).

Depending on the types of studies that participants have conducted and the analyses their data require, the scope of the presentations can be expanded (more on statistical tests, for example, or more on analysis of qualitative data), or the sequence changed (Module 23 may be presented before module 22 if participants have mainly qualitative data).

- Usually the *first half of the workshop* (one week) is devoted to the finalisation of data processing and to data analysis. All modules related to analysis (21-31) are presented during this week.
- Timing of these presentations has to be done carefully. Modules 21-24 can be presented before data processing has been completed. Modules 25 and 26 could be presented when groups are about to finish data processing and have finished some basic tables. Module 27, preparing for statistical analysis (measures of dispersion, distribution and sample variation) can then follow as well.
- Only when participants are well underway with the preparation of cross-tabulations, should the modules that present various statistical tests be presented.
- The *second half of the workshop* concentrates on report writing, drafting of recommendations, and presentation and discussion in plenary of the main findings and recommendations arising from the studies. In this week there are usually only two presentations: on report writing (Module 32) and on the dissemination, communication and utilisation of research findings (Module 33). The last module is best presented just before the participants draft the summary of findings and the recommendations of their studies.

An example of a schedule for a 2-week course on data analysis and report writing is presented on the following pages.

If the level of participants is high and if the data have been satisfactorily processed before reconvening for the data analysis and report writing workshop, it may be possible to finish the report including a draft Plan of Action within two weeks. Otherwise, the finishing touches will have to be accomplished afterwards. Some support of a facilitator, either live or by computer, will then be required.

EXAMPLE OF A COURSE SCHEDULE

(as used in southern Africa)

Designing and Conducting HSR Projects: Data Analysis and Report Writing

Date/Time	Session	Responsible Person(s)
Monday		
08.00 – 08.30	Opening remarks	Course Coordinator
08.30 – 09.15	Presentation and discussion of preliminary results	Group 1
09.15 – 10.00	Presentation and discussion of preliminary results	Group 2
10.00 – 10.30	Tea	
10.30 – 11.15	Presentation and discussion of preliminary results	Group 3
11.15 – 12.00	Presentation and discussion of preliminary results	Group 4
12.00 – 12.30	Module 21: Orientation to the workshop on data analysis and report writing	Facilitator
12.30 – 14.00	Lunch	
14.00 – 15.30	Group work	
15.30 – 16.00	Tea	
16.00 – 17.00	Module 22: Description of variables	Facilitator
17.00 – 18.00	Group work	
Tuesday		
08.00 – 09.00	Module 23: Analysis of qualitative data	Facilitator
09.00 – 13.00	Group work (including tea)	
13.00 – 14.00	Lunch	
14.00 – 15.00	Module 24: Cross-tabulation of quantitative data	Facilitator
15.00 – 18.00	Group work (including tea)	

Wednesday

08.00 – 13.00	Group work (including tea)	
13.00 – 14.00	Lunch	
14.00 – 15.00	Optional: Presentations of main results of group work: revised objectives, main cross-tables, results of qualitative analysis	All 4 Groups
15.00 – 16.00	Module 25: Measures of association based on risk (incidence, risk, relative risk, odds ration)	Facilitator
16.00 – 18.00	Group work (including tea)	

Thursday

08.00 – 09.00	Module 26: Dealing with confounding	Facilitator
09.00 – 13.00	Group work (including tea)	
13.00 – 14.00	Lunch	
14.00 – 15.00	Module 27: Preparing for statistical analysis (Measures of dispersion, normal distribution and sample variation)	Facilitator
15.00 – 18.00	Group work (including tea)	

Friday

08.00 – 09.00	Module 28: Choosing a significance test (20 min) followed by Module 29, Part I and II (t-test)	Facilitator
09.00 – 13.00	Group work (including tea)	
13.00 - 14.00	Lunch	
14.00 - 15.00	Module 29 part III (Chi square test)	Facilitator
15.00 - 18.00	Group work (including tea)	

Saturday

08.00 – 13.00	Group work (including tea) Modules 30 or 31 (if required)	
---------------	---	--

Sunday

Free

Monday

08.00 – 09.00

Rest of day

Module 32: Report writing

Group work

Facilitator

Tuesday

Whole day

Group work

Wednesday

08.00 – 08.30

Rest of day

Module 33: Promoting the dissemination, communication and utilisation of findings

Group work

Facilitator

Thursday

08.00 – 13.00

13.00 – 14.00

14.00 – 17.30 +

Group work

Lunch

Group work; Preparation of presentation of research results, recommendations, and preliminary Plan of Action

All groups and facilitators

Friday

08.00 – 13.00

13.00 – 14.00

14.00 – 17.00

17.00 – 17.30

Group work to finalise reports and presentations

Lunch, with invited guests (Interested health and research managers of the MOH and university/research institutions)**Presentations of summary of findings, recommendations and tentative Plan of Action** by the four groups, followed by discussion (3/4 hour per group, with tea break in between)**Evaluation** of the HSR Training Course**Saturday** (if necessary)

08.00 – 13.00

Finishing touches to reports

This page intentionally left blank

**Designing and Conducting Health Systems Research Projects
Part II: Data Analysis and Report Writing**

Module 21

**ORIENTATION TO THE WORKSHOP
ON DATA ANALYSIS AND REPORT WRITING**

Module 21: ORIENTATION TO THE WORKSHOP ON DATA ANALYSIS AND REPORT WRITING

OBJECTIVES OF THE WORKSHOP

At the end of this workshop you should be able to:

1. **Identify** and define the basic concepts and procedures required for data analysis and interpretation.
2. **Analyse** and interpret the data collected for the research project which you developed during the first workshop and draw conclusions related to the objectives of your study.
3. **Write** a clear and concise research report and a summary of the major findings and recommendations for each of the different parties interested in the results.
4. **Present** the major findings and the recommendations of your study to policy-makers managers and to the subjects of your research together with them to finalise the recommendations.
5. **Prepare** a plan of action for the dissemination, communication and utilisation of the findings and (if required) make recommendations for additional research.

I. Review of the field experience

II. Introduction to the workshop

III. Tasks to be completed during the workshop

1. Review and finalisation of data processing
2. Data analysis
3. Report writing
4. Presentation of summary of findings and recommendations
5. Drafting a plan for the implementation of the research results

I. REVIEW OF FIELD EXPERIENCES

Implementing your planned project proposal must have been a big challenge to you. No doubt you met a number of unexpected obstacles as you became involved in your fieldwork but you will have experienced successes as well. If everything went according to plan you have collected your data; you have processed a large part of it, if not all; you have completed some of your analysis; and you have written a preliminary report on the experiences and results of your fieldwork. Your practical experiences in conducting the project are invaluable. Sharing these experiences with others in this workshop will be a very useful exercise, as both your problems and successes can provide valuable lessons for the future.

Before providing an overview of the focus and activities of this workshop, we would like to spend some time listening to the experiences of each of the research groups.

EXERCISE: Presentations of field experiences

Present the preliminary report that your group prepared at the end of your field experience, following the guidelines given in **Module 20**. Be prepared to answer any questions other participants or facilitators may pose at the end of your presentation.

Each group will have approximately 10 – 15 minutes for its presentation and then some time for questions and discussion.

II. INTRODUCTION TO THIS WORKSHOP

This workshop is a follow-up of the workshop in which you developed your proposal. Now you have the major task ahead of fully analysing the data you brought with you from the field and writing your research report. The report should contain feasible and useful recommendations, based on the findings of your study concerning how to solve the problem investigated.

As in the first workshop there will be presentations, group work sessions and a few plenaries. In this workshop, however, group work will take up most of the time. The presentations will be concentrated in the first week, which will be devoted to data analysis. The second week will be fully reserved for report writing, with only two presentations to guide you. Toward the end of that week an important plenary is planned in which each group will present a summary of its main findings and recommendations. A selected group of policy-makers and health managers who requested the study or have a direct interest in the topic and some interested researchers will be invited to comment on your presentation during that plenary.

The modules for this workshop cover several major tasks, which are schematically presented in the diagram on the next page of this module. This diagram is presented again at the beginning of each subsequent module, to indicate which task is the focus of the presentation. We will now briefly introduce each of these tasks.

Steps in data analysis and report writing

Questions you must ask	Steps you will take*	Important elements of each step
<p>What data have been collected for each research objective? Are data complete, accurate?</p>	<div style="border: 1px solid black; padding: 5px; background-color: #e0f2f1; width: fit-content; margin: 0 auto;">Prepare data for analysis</div>	<p>Review field experience Make an inventory of data for each objective/study population Sort data and check quality Check computer outputs (21)</p>
<p>What do the data look like? How can the data be summarised for easy analysis?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Summarise data and describe variables/identify new variables</div>	<p>Frequency tables, figures, means, proportions, descriptive cross-tabulations, (<i>quantitative data</i>) (22, 24); Coding, listing, summarising data in compilation sheets, matrices, flow charts, diagrams and narratives (<i>qualitative data</i>) (23)</p>
<p>How can the associations between variables be determined?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Analyse associations</div>	<p>Analytic cross-tables (24) Measures of association based on risk(25) Dealing with confounders (26)</p>
	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Prepare for statistical analysis</div>	<p>Measures of dispersion, Normal distribution and Sampling variation (27)</p>
<p>Do we measure differences or associations between variables?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Determine the types of statistical analysis</div>	<p>Choosing significance tests (28)</p>
<p>How can differences between groups be determined?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Analyse unpaired and paired observations</div>	<p>t-test, chi-square test (29) ** paired t-test, ** McNemar's chi-square test (30)</p>
<p>How can the associations between numeric variables be determined?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Implement measures of association</div>	<p>** Scatter diagram, ** Regression line and ** Correlation coefficient (31)</p>
<p>How should the report be written?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Write the report and formulate recommendations</div>	<p>Prepare outline for report Present and interpret data Draft and redraft Discuss and summarise conclusions Formulate recommendations (32)</p>
<p>How should the findings and recommendations be communicated, disseminated and used?</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Present summaries and draft for implementation of recommendations</div>	<p>Discuss summaries and plan for implementation with all stakeholders (33)</p>

* These steps need not be in the sequence in this diagram. **The sequence may be adjusted according to the needs of the research teams.**

** These elements are optional and may be omitted if not relevant for research teams.

I. TASKS TO BE COMPLETED DURING THE WORKSHOP

1. Review and finalisation of data processing

Although we trust that all groups have put great effort in the processing of their data, some adjustment and elaboration may be required. This is normal in research: each new step forward may be followed by half a step backward. Even in an advanced stage of data analysis you may still have to regroup and reprocess some of your data.

Before beginning data analysis, it is extremely important to check whether data processing has been carried out in such a way that information:

- is easy to handle; and
- has been checked for mistakes that may have crept in during data collection.

You therefore have to ask yourself the following questions:

- **Have the data been sorted appropriately?** Have questionnaires and checklists been numbered in the most convenient way? Can major categories of informants (e.g., m/f, cases and controls) be clearly distinguished to facilitate comparison on relevant variables, as required by your research objectives?
- **Have quality checks been performed** on all data for completeness and consistency of information? Look at **Module 13** for measures to be taken in case of incompleteness and/or inconsistencies.
- **Have all data been entered in the computer or, if using master sheets, have all data been filled?** Do the total number of responses match with the total number of respondents for each variable? If not: have some unknowns or missing data been overlooked?
- **Has all qualitative data been categorised** as far as possible? If applicable, has coding been completed? (See **Module 13** for post-categorising of open-ended questions.) Have FGDs been carefully read and ordered according to the discussion topics? Have particularly illustrative parts in relation to the research objectives or research questions been highlighted? Since for qualitative data the collection, ordering, summarising and analysis are, in principle, intertwined (see module 10C), you will already have coded and interpreted a large part of your qualitative data. **Module 23** will take the analysis of qualitative data up in detail.
- If you used the computer to process your data, **check the frequency counts for each variable** in the questionnaire. Also check the computer cross-tabulations. Details on how to do this are given in **Annex 21.1**.

Before reviewing the data processing procedures it is strongly advised that you make an **INVENTORY** of all data available for each **OBJECTIVE**. This is especially important if the data required has been collected using different data collection tools.

Example:

Data sources for Objective 3: 'Detection of weaknesses in the functioning of MCH services, explaining low utilisation of delivery care:'

- Questionnaire for mothers, Questions 12, 15 - 19, 23
- Focus group discussion with health staff, topics 3 and 4
- Observations included in checklists

Such an inventory will help you to better organise data analysis and, later, report writing.

2. Data analysis

When beginning data analysis, we should consider which of our data are quantitative and which are qualitative.

Quantitative data

Quantitative data are expressed in numbers and they are usually presented in frequency tables. From your data master sheets you can easily derive totals for each variable/question, count the number of different answers obtained and present the information in frequency tables. (See **Module 22**.)

When analysing quantitative data it is important to consider the aim of your study. Is it to:

- **Describe variables?**

For example: the distribution of teenage pregnancies in a certain population

- **Look for differences between groups?**

For example: differences between old settlers and newcomers in a certain area, with respect to income or health status

- **Determine associations between variables?**

For example: the association between work satisfaction of nurses and the number of staff meetings over the past year

Cross-tabulations are an important tool to summarise and analyse these data (**Module 24**), though there are other possibilities (see **Modules 22, 25** and **31**).

After frequency distributions and different types of cross-tabulations have been made, the type of statistical analysis required has to be selected in order to determine whether the differences and associations found are significant or just a consequence of chance. The selection of appropriate significance tests is elaborated in **Module 28**. In **Modules 29-31** some more advanced statistical concepts for the analysis of quantitative data will be introduced.

The most common significance tests are:

- Student's t-test and the chi-square test to determine differences between groups if observations are unpaired (**Module 29**).
- The paired t-test and McNemar's χ^2 (chi-square) test to determine differences between groups for paired observations (**Module 30**).

For measuring associations between variables, the concepts of Odds Ratio (**Module 26**) and Regression and Correlation (**Module 31**) will be introduced.

Throughout the process of data analysis it is important to keep in mind that *our findings should provide answers to our research questions and thus meet our research objectives. We will eventually want to draw conclusions and make recommendations for action, based on these findings.*

Qualitative data

You will remember that we may obtain qualitative data through:

- **open-ended questions**, not pre-categorised, in questionnaires or interview schedules which also collect quantifiable data;
- **loosely structured interviews** with predominantly open-ended questions, directed at key informants (individuals) or small groups;
- **focus group discussions** on selected issues, with lists of points to guide the discussions;
- **observations** describing individual or group behaviour;
- **diaries, essays**, and any information that originates from **projective methods** (e.g., unfinished sentences, stories with a gap, free associations of informants with pictures or films shown).

As you will remember from the exercise you did in **Module 13**, the answers to open-ended questions may be:

- **listed**
- **categorised** (based on your research objectives and common sense, combining the answers that belong together in some 4 to 6 categories, rarely more)
- **coded/ labelled**
- **interpreted** per category for content, depending for what purpose you need the data
- **inserted**, using these codes, **in your master sheets, or in the computer**
- **counted**, like other quantitative data

You may have already processed some clear-cut open questions. We will discuss these procedures again in-depth in **Module 23** in case you experienced some problems.

Note:

The major characteristic of analysis of qualitative data is that we analyse it **IN WORDS**, rather than in numbers.

Qualitative data from other sources than open-ended questions require more elaborate coding and compilation techniques. Often it is useful to summarise qualitative data in compilation sheets, diagrams, flow-charts, or matrices which help us in our analysis. **Module 23** will deal with the analysis of such qualitative data in more detail.

3. Report writing

You will be expected to go home with a completed report of your research. It will have the following components:

1. An INTRODUCTION, covering the statement of the problem, some relevant contextual data and literature review.
2. OBJECTIVES
3. A METHODOLOGY section with information on when, where and how you have collected your data, how you have analysed the data; and possible weaknesses in the collection and analysis.
4. FINDINGS
5. DISCUSSION
6. CONCLUSIONS AND RECOMMENDATIONS

The last three sections, which will form the bulk of your report, will be discussed in detail in **Module 32**. The first three sections can be revised and summarised from the relevant sections in your research proposal.

4. Presentation of summary of findings and recommendations

Since an important goal of your research is that appropriate action will be undertaken based on the results of your study, it is important that all parties concerned get an opportunity to discuss findings and recommendations before the report is finalised. You may wish to include policy-makers, health managers, staff and community members or even the media in such a discussion. **Module 33** provides some guidelines on how to organise meetings for this purpose.

5. Drafting a plan for the implementation of the research results

You drafted a plan for utilisation and dissemination of results during the previous workshop (**Module 17**). At the end of the present workshop this plan will be reviewed and developed in more detail, including all parties concerned in the planning for implementation of the recommendations that resulted from your study (**Module 33**).

GROUP WORK (Time flexible, depending on the research topics and state of data processing)

- Reconsider the objectives from your research proposal and **list the different data sources for each objective** (questionnaires, records, focus group discussions etc.).

NB: If you discover that you have collected more data to explain your research problem than your objectives require, you may review your objectives or add one or two additional ones. However, if you collected less data, don't drop objectives which you could not meet but explain why you couldn't.

- Verify whether all data has been checked for completeness, consistency, and proper coding. If not, do so.
- Determine whether different master sheets have been prepared for different study populations or for different categories of informants you would like to compare to each other. This will facilitate analysis.

You may also mark the questionnaires of sub-groups with different colours so that you can easily refer back to the raw data to check certain questions.

- Check whether the master sheets have been completed and whether the number of responses for each variable agrees with the number of respondents.
- Determine whether all data that should have been entered in the computer have indeed been entered and cleaned. (See Annex 21.1.)
- Check whether qualitative data were categorised and summarised in the field. If not, read and order the data, putting discussion topic numbers and additional key words and comments in the margins. In **Module 23** further advice will be provided.
- As you list your different data sources by objective, check whether you have recorded all your relevant observations.

Annex 21.1: Computer output

The hard copy printed out by the computer is the result of the commands used in the computer programmes to analyse the available data. The accuracy of the information printed out is therefore dependent on:

- the data that was entered
- the programmes that were used

The saying, 'garbage in, garbage out', is very apt for computer processing. It is the combined responsibility of the research team and computer specialist to ensure that the information printed out is accurate.

Types of computer printouts

1. List of data

This is a list of the data that was entered into the computer. This printout is helpful if you need to make corrections on the existing data while in the process of validating it.

2. Frequency count

This gives a count (and percentage) of each variable in the questionnaire. See the sample given of a frequency count and note how it relates to the questionnaires.

To ensure that the programmes are correct, the computer specialist must be familiar with the format of the questionnaire used and the process of data collection.

For example:

Was stay in hospital appropriate? (n = 3306)

Yes	No	Don't know	Total
1937	1369	0	3306

Was admission appropriate? (n = 3306)

Yes	No	Don't know	Total
634	719	16	1369

In this example, the total responses for 'yes', 'no' and 'don't know' for the question 'Was the stay in hospital appropriate?' is equal to the total study cases. The total responses for the question, 'Was admission appropriate' are only 1369. The computer specialist must be aware that this question was only asked to the informants who were dissatisfied with their stay in hospital, otherwise the computer print will indicate that 1937 answers are missing instead of not applicable (NA).

A frequency count should be obtained for every question in the questionnaire. Use the frequency count to ensure that:

- the total number of responses in each question is correct (i.e., it should tally with the sample size of persons being asked the question);
- all codes are relevant to the question. **For example**, there should be no codes 3-8 in a question that has only two possible responses (e.g. sex: M or F) and a code for 'unknown' (unknown is usually given a code 9).

3. Cross tabulation

The next commonest computer output is a **cross-tabulation**. This is a table showing the number of subjects who have two (or more) of the variables studied.

Example:

	Male	Female	Total
ill			
not ill			
total			

Before using it, check the cross-tabulation for the following:

- The grand total in the table should correspond to the number of subjects in the sample
- Column and row totals should correspond to the frequency counts for each variable (i.e., the number of males and females should correspond to the respective frequency counts)
- Similarly, numbers 'ill' and 'not ill' should correspond to that frequency count. If these do not correspond, there is probably an error in the programme. Consult your computer specialist.
- If there is a statement in the computer printout showing 'missing cases' it means either:
 - there is a wrong code in the data entry (e.g., code 4 when only 1,2 or 9 is possible), or
 - the categories you have specified are not comprehensive.

For example:

The questionnaire allowed for 'unknown' but the computer programme did not. Therefore all cases 'unknown' would appear as 'missing cases'.

Marital status in the questionnaire allowed for 'married, single, divorced, widowed'. However, the computer programme specified only 'married, single, divorced'. All widowed persons would be missing.

If the age categories are 10 to 14, 15 to 19 but the programmer accidentally programmed the categories as 10 to 13, 15 to 19, all subjects aged 14 would be missing.

Module 21: ORIENTATION TO THE WORKSHOP ON DATA ANALYSIS AND REPORT WRITING

Timing and teaching methods

3 hours	Presentations of field experiences
$\frac{3}{4}$ hour	Introduction and discussion
1 hour +	Group work (duration depending on research topics and state of data processing)

Introduction and discussion

- Spend the first part of the introductory session on the participants' reports of their field experiences. If all the groups have come prepared to present their preliminary reports, the session can begin directly with this activity. However, if groups still need some time to prepare for their presentations, time should be arranged for this, either before or at the beginning of this first session.
- The introduction to the workshop should clearly stress that there are different tasks to complete, of which data analysis and reporting of the findings will be the most time consuming. It should be clear to the participants, however, that the preparation of recommendations and their implementation is the ultimate aim of their research projects. You might ask the participants for *suggestions* concerning *which policy makers and managers should be invited* for the presentation and discussion of their research findings and recommendations at the end of the workshop.
- When presenting the diagram consider using different, overlapping transparency sheets.
- Adjust the presentation to the level and interests of the participants. Refresh their memory with examples of the processing of open-ended questions and by explaining the difference between descriptive studies, comparison of groups, studies looking for differences between groups, and studies determining associations between variables - **preferably with examples from their own research**.
- Do not frighten groups that have little statistical experience with details on tests at this stage. Merely state that each type of study requires different tests.
- Stress the importance of listing all the data available for each objective, including qualitative data. As the workshop proceeds there will be so much emphasis on the preparation of tables that participants will tend to forget valuable observations and information obtained from key informants. The facilitator should ask the participants to record this information now (if not already done) and include it in the list of data available for each objective. Check, when the report is being written, that it has been analysed.

Group work

- Read the group work directions along with the group members. Let them re-examine their objectives, list the data available for each objective, and discuss whether the objectives are specific enough to cover all relevant data collected. Sometimes objectives have to be split up, rephrased, added, or their order changed to facilitate analysis. **Never** allow the group to omit an objective without an explanation (in the methodology section) concerning why it could not be met.
- Examine, with the group members, all available data for completeness, mistakes, etc. Make sure that separate master sheets have been prepared for different study populations or for different subgroups that will be compared, or that the data on different subgroups can be easily retrieved from the computer.

Take extra time, as a facilitator, to internalise all data available, to identify possible weaknesses, and to consider various possibilities for analysis. Unless you do this at the onset of the workshop it will be difficult to guide the groups efficiently so that they will obtain optimal results from the data they have collected.

- Group members may work in sub-groups to finalise the data processing, but make sure that you discuss problems and progress with the group as a whole at regular intervals.

This page intentionally left blank

**Designing and Conducting Health Systems Research Projects
Part II: Data Analysis and Report Writing**

Module 22

DESCRIPTION OF VARIABLES

Steps in data analysis and report writing

Questions you must ask	Steps you will take*	Important elements of each step
What data have been collected for each research objective? Are data complete, accurate?	Prepare data for analysis	Review field experience Make an inventory of data for each objective/study population Sort data and check quality Check computer outputs (21)
What do the data look like? How can the data be summarised for easy analysis?	Summarise data and describe variables/identify new variables	Frequency tables, figures, means, proportions, descriptive cross-tabulations, (<i>quantitative data</i>) (22, 24); Coding, listing, summarising data in compilation sheets, matrices, flow charts, diagrams and narratives (<i>qualitative data</i>) (23)
How can the associations between variables be determined?	Analyse associations	Analytic cross-tables (24) Measures of association based on risk(25) Dealing with confounders (26)
	Prepare for statistical analysis	Measures of dispersion, Normal distribution and Sampling variation (27)
Do we measure differences or associations between variables?	Determine the types of statistical analysis	Choosing significance tests (28)
How can differences between groups be determined?	Analyse unpaired and paired observations	t-test, chi-square test (29) ** paired t-test, ** McNemar's chi-square test (30)
How can the associations between numeric variables be determined?	Implement measures of association	** Scatter diagram, ** Regression line and ** Correlation coefficient (31)
How should the report be written?	Write the report and formulate recommendations	Prepare outline for report Present and interpret data Draft and redraft Discuss and summarise conclusions Formulate recommendations (32)
How should the findings and recommendations be communicated, disseminated and used?	Present summaries and draft for implementation of recommendations	Discuss summaries and plan for implementation with all stakeholders (33)

* These steps need not be in the sequence in this diagram. The sequence may be adjusted according to the needs of the research teams.

** These elements are optional and may be omitted if not relevant for research teams

Module 22: DESCRIPTION OF VARIABLES

OBJECTIVES

At the end of this session you should be able to:

1. **Describe** data in terms of frequency distributions, percentages, and proportions.
2. **Use** figures to present data.
3. **Explain** the difference between mean, median and mode.
4. **Calculate** the frequencies, percentages, proportions, ratios, rates, means, medians, and modes for the major variables in your study that require such calculations.
5. **Identify** other independent variables (in addition to the ones identified during the first workshops), if any, that are necessary in the analysis of your data.

I . Introduction

II. Frequency distributions

III. Percentages, proportions, ratios, and rates

IV. Figures

V. Measures of central tendency

I. INTRODUCTION

When you selected the variables for your study in **Module 8**, you did so with the assumption that they either would help to define your problem (dependent variables) and its different components or that they were contributory factors to your problem (independent variables). The purpose of data analysis is to identify whether these assumptions were correct or not, and to highlight possible new views on the problem under study. **The ultimate purpose of analysis is to answer the research questions outlined in the objectives** with your data.

First, before we look at how variables may be affecting one another, we need to summarise the information obtained on each variable in simple tabular form or in a figure.

Some of the variables may have produced numerical data, while other variables produced categorical data. In analysing our data, it is important first of all to determine the type of data that we are dealing with. This is crucial because the type of data used largely determines the type of statistical techniques that should be used to test whether the results of the study are significant.

Categorical data

There are two types of categorical data: they are nominal or ordinal (see **Module 8**).

In **NOMINAL DATA**, the variables are divided into a number of named categories. These categories, however, cannot be ordered one above another (as they are not greater or lesser than each other).

For example:

NOMINAL DATA	CATEGORIES
Sex	male, female
Marital status	single, married, widowed, separated/divorced

In **ORDINAL DATA**, the variables are also divided into a number of categories, but these can be ordered one above another, from lowest to highest or vice versa

For example:

ORDINAL DATA	CATEGORIES
Level of knowledge	good, average, poor
Opinion on a statement	fully agree, agree, doubt, disagree, totally disagree

Numerical data

We speak of NUMERICAL DATA if they are expressed in numbers

There are two types of numerical data: they are discrete or continuous.

DISCRETE DATA are a distinct series of numbers.

For example:

DISCRETE DATA	VALUES
Number of motor vehicle accidents	0, 1, 2, 6, 19, etc.
Number of clinic visits	2, 4, 10, 0, 3, etc.
Number of pregnancies per woman	2, 12, 5, 0, 5, 4, etc.

CONTINUOUS DATA come from variables that can be measured with greater precision, depending on the accuracy of the measuring instrument, and each value can increase or decrease without limit.

For example:

CONTINUOUS DATA	VALUES
Height (to 2 decimal point)	12.12, 9.95, 45.13, 6.99, 28.78, etc,
Temperature (in degree Celsius)	37.5, 37.8, 39.2, 40.1, 36.9, etc,
Age (at the last birthday)	50, 45, 12, 78, 25, 16, 61, 90

Numerical data can be presented as:

- Frequency distributions
- Percentages, proportions, ratios and rates
- Figures
- Measures of central tendency

We will now discuss these operations one after each other for both categorical and numerical data.

II. FREQUENCY DISTRIBUTIONS

A **FREQUENCY DISTRIBUTION** is a description of data presented in tabular form so the data will be more manageable. It gives the frequency with which a particular value appears in the data.

In your research project you will have done already straight frequency counts for all variables in your data master sheets by counting the number of responses in each category. We will now briefly summarise some important points.

1. **Categorical data** may have very simple categories.

Example 1:

To identify what family planning methods were used by teenagers in Kweneng, West Botswana, teenagers were asked what method they were using.

The results are presented in the following **frequency distribution**:

Method	Number
Abstinence	14
Condoms	47
Injectables	1
Norplant	1
Pill	35
None	307
Total	405

These data are **NOMINAL**. A frequency distribution is calculated by simply totalling the number of responses in each category.

You should always check that the total number of responses agrees or tallies with the number of subjects (respondents). If necessary, there should be a category for missing answers.

We usually express frequency distributions in percentages (see Part III of this Module). By looking at the frequency distribution above you can conclude that roughly 75% or three out of four of the teenagers are not using family planning. For those who are using family planning methods, condoms and pills are the most commonly used methods.

Example 2:

Health personnel from 148 different rural health institutions were asked the following question: How often have you run out of drugs for the treatment of malaria in the past two years? This was a closed question with the following possible answers: never, 1 to 2 times (rarely), 3 to 5 times (occasionally), more than 5 times (frequently).

The number of responses in each category was totalled to give the following frequency distribution:

Categories	Number
Never	47
Rarely	71
Occasionally	24
Frequently	6
Total	148

In this example, the data are **ORDINAL**. The ordering of the categories is important as each category from top to bottom indicates increasing severity of the problem.

The frequency distribution results indicate that most clinics rarely experience shortages of anti-malarial drugs, but that it is an occasional problem in about one sixth of the clinics and a severe problem in a few.

2. Numerical data

Procedures for making frequency distributions of numerical data are very similar to those for categorical data, except that now the data have to be grouped in categories. The steps involved in making a frequency distribution are as follows:

1. Select groups for grouping the data.
2. Count the number of measurements in each group.
3. Add up and check the results.

When grouping data, the way the groups are selected can affect what the results are going to look like. There is little substitute for common sense here, but it may be necessary to change the grouping if you suspect the information is being hidden by a poor selection of the groups.

Example 3:

Health centres of District X are submitting numbers of malaria cases and you wish to summarise them. Compare the daily and weekly summaries of the same data as presented in **Table 22.1**:

Both daily and weekly data show an increasing amount of malaria, but the improving situation shown in days 19, 20 and 21 is not reflected in the weekly summary. It would therefore be better to use the daily data if you want to indicate when exactly the numbers of reported malaria cases started going down.

Table 22.1: Daily and weekly summaries of malaria cases in health centres in District X

Day 1	9 cases	Week 1	88 cases
Day 2	12		
Day 3	11		
Day 4	13		
Day 5	14		
Day 6	13		
Day 7	16		
Day 8	16 cases	Week 2	131 cases
Day 9	16		
Day 10	18		
Day 11	19		
Day 12	16		
Day 13	21		
Day 14	25		
Day 15	28 cases	Week 3	168 cases
Day 16	28		
Day 17	28		
Day 18	32		
Day 19	21		
Day 20	19		
Day 21	12		

When grouping data the following rules are important:

- The groups must not overlap, otherwise there is confusion concerning in which group a measurement belongs.
- There must be continuity from one group to the next, which means that there must be no gaps. Otherwise some measurements may not fit in a group.
- The groups must range from the lowest measurement to the highest measurement so that all of the measurements have a group to which they can be assigned.
- The groups should normally be of an equal width, so that the counts in different groups can easily be compared.

Sometimes, however, it is valid to choose groups that are of different widths, for example if you are interested in specific age groups (e.g., less than 1 year, 1 to 4 years, 5 to 14 years).

When you start summarising data it is better to make too many groups than too few. This is because during data analysis you can combine groups to form new categories without having to go through all your data again, whereas if you have too few groups you have to go back to your raw data to make more groups.

A larger number of groups will generally give a more precise picture, but when using too many groups one can lose the overview.

As a general rule choose round numbers for the lower values of the group limits.

For example: 1.00-9.99, 10.00-19.99, 20.00-29.99, or:
0-4; 5-9, 10-14, etc.

III. PERCENTAGES, PROPORTIONS, RATIOS, AND RATES

1. Percentages

Instead of presenting data in frequency tables using absolute numbers it is often better to calculate percentages.

A **PERCENTAGE** is the number of units in the sample with a certain characteristic, divided by the total number of units in the sample and multiplied by 100.

Percentages may also be called **RELATIVE FREQUENCIES**. Percentages *standardise* the data, which means that they make it easier to compare them with similar data obtained in another sample of different size or origin.

Example 4:

82 clinics in one district were asked to submit the number of patients treated for malaria in one month. The researchers presented both the frequency distribution and percentages (or relative frequencies):

Table 22.2: Distribution of clinics according to number of patients treated for malaria in one month

Number of patients	Number of clinics	Relative frequency
0 to 19	25	31%
20 to 39	3	4%
40 to 59	5	6%
60 to 79	11	14%
80 to 99	19	24%
100 to 119	10	12%
120 to 139	4	5%
140 to 159	3	4%
Total	80	100%

Note:

Usually you do not include missing data in the calculation of percentages.

The frequency of responses in each group is calculated as the percentage of those study elements for which you obtained data (or, if a question is being asked to interviewees, the percentage of those interviewees who answered the question).

However, the number of missing data (e.g., people who did not respond to a question) is a useful indication of the adequacy of your data collection. Therefore this number should be mentioned, for example as a note to your table. (See **Table 22.2.**)

Remember that 'don't know' is a special category that should NOT be counted as missing data. If applicable, 'don't know' should appear as a category in the table.

One should be cautious when calculating and interpreting percentages if the total number is small, because one unit more or less would make a big difference in terms of percentages. As a general rule, percentages should not be used when the total is less than 30.

Therefore it is recommended that the number of observations or total cases studied should always be given together with the percentage.

2. Proportions

Sometimes relative frequencies are expressed in proportions instead of percentages.

A PROPORTION is a numerical expression that compares one part of the study units to the whole; A proportion can be expressed as a FRACTION or in DECIMALS.

Example 5:

Out of a total of 55 patients attending a clinic on a specific day 22 are males and 33 are females. We may say that the proportion of males is $22/55$ or $2/5$, which is equivalent to 0.40.

Note that when a proportion expressed in decimals is multiplied by 100, the value obtained is a percentage. In the example, 0.40 is equivalent to 40%.

3. Ratios

A RATIO is a numerical expression that indicates the relationship in quantity, amount or size between two or more parts.

In **Example 5** above the ratio of males to females is 22:33, or 2:3.

4. Rates

A RATE is the quantity, amount or degree of a disease or event measured over a specified period of time

Commonly used rates in the health sector are:

- Birth Rate = The number of live births per 1000 population over a period of one year
- Death Rate = The number of deaths per 1000 population over a period of one year

- Infant Mortality Rate (IMR) = The number of deaths of infants under one year deaths of age per 1000 live births over a period of one year
- Maternal Mortality Rate (MMR) = The number of maternal pregnancy-related in one year per 100,000 total births in the same year
- Incidence Rate = The number of new cases per population over a specific period of time (usually a year)
- Prevalence Rate = The number of existing cases per population over a specific period of time (usually a year)

IV. FIGURES

If your report contains many descriptive tables, it may be more readable if you present the most important ones in figures.

The most frequently used figures for presenting data include:

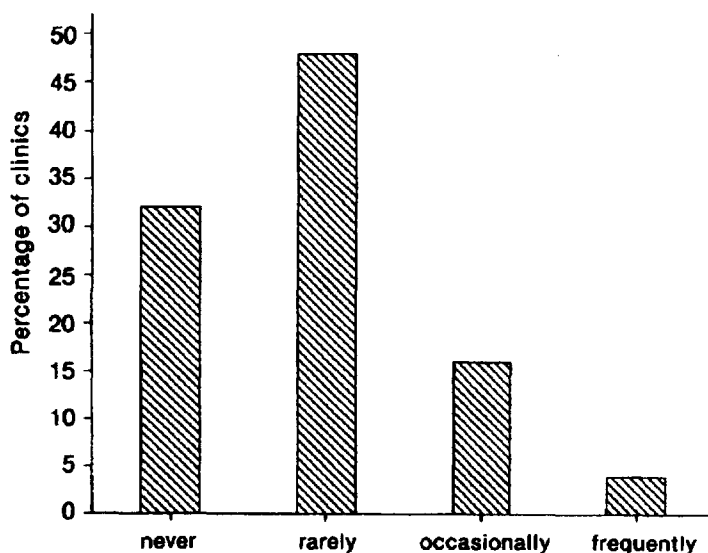
- Bar charts
 - Pie charts
- } for categorical data
- Histograms
 - Line graphs
 - Scatter diagrams
 - Maps
- } for numerical data

We will now look at example of the above-mentioned figures that can be used for presenting data.

1. Bar chart

The data from **Example 2** can be presented in a bar chart, using either absolute frequencies or relative frequencies/percentages (see **Figure 22.1**).

Figure 22.1: Relative frequency of shortage of anti-malaria drugs in rural health institutions (n = 148)

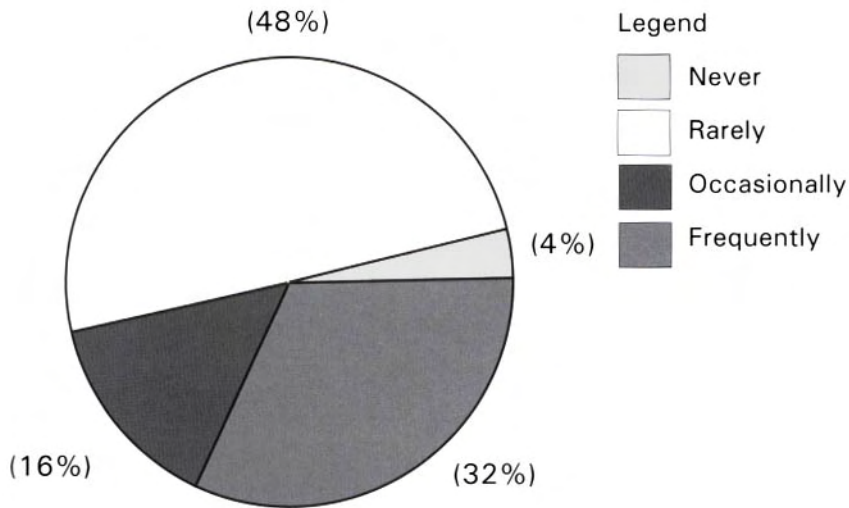


Note that the sample size must be indicated if you present the data in percentages.

2. Pie charts

A pie chart can be used for the same set of data, providing the reader with a quick overview of the data presented in a different form. A pie chart illustrates the relative frequency of a number of items. All the segments of the pie chart should add up to 100%.

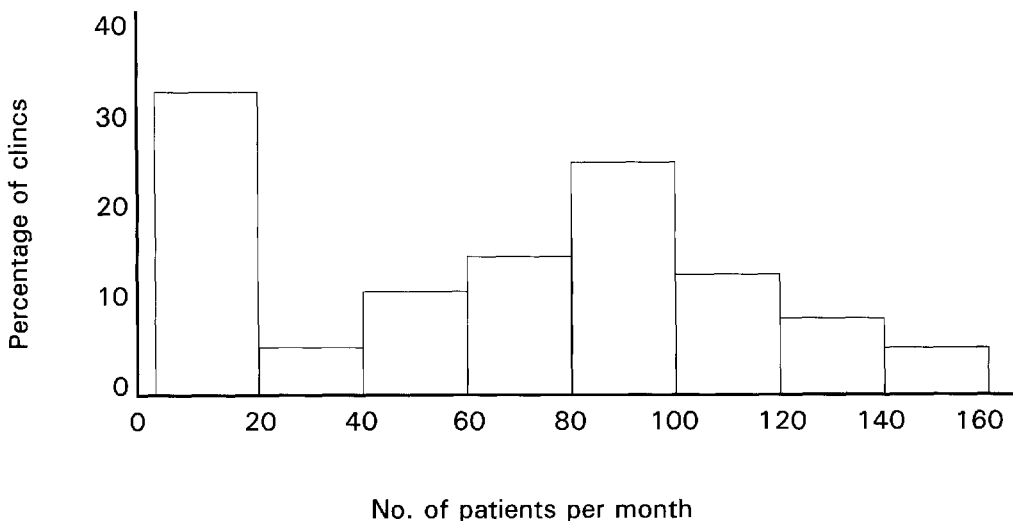
Figure 22.2: Relative frequency of shortage of anti-malaria drugs in rural health institutions (n = 148)



3. Histograms

Numerical data are often presented in histograms, which are very similar to the bar charts which are used for categorical data. An important difference however is that in a histogram the 'bars' are connected (as long as there is no gap between the data), whereas in a bar chart the bars are not connected, as the different categories are distinct entities. The data of **Example 4** is presented as a histogram in **Figure 22.3**.

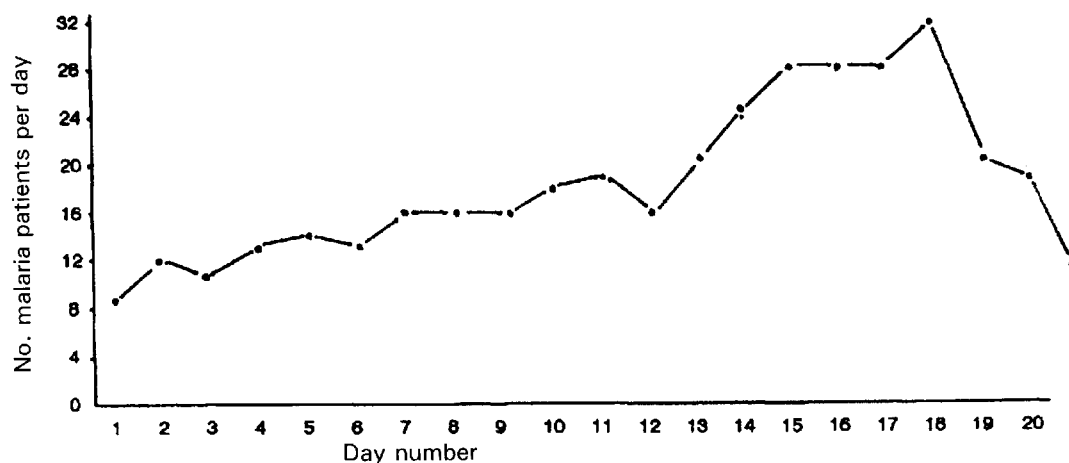
Figure 22.3: Percentage of clinics treating different numbers of malaria patients in one month (n = 80).



4. Line graphs

A line graph is particularly useful for numerical data if you wish to show a trend over time. The data from **Example 3** can be presented as a line graph as in **Figure 22.4**.

Figure 22.4: Daily number of malaria patients at the health centres in District X

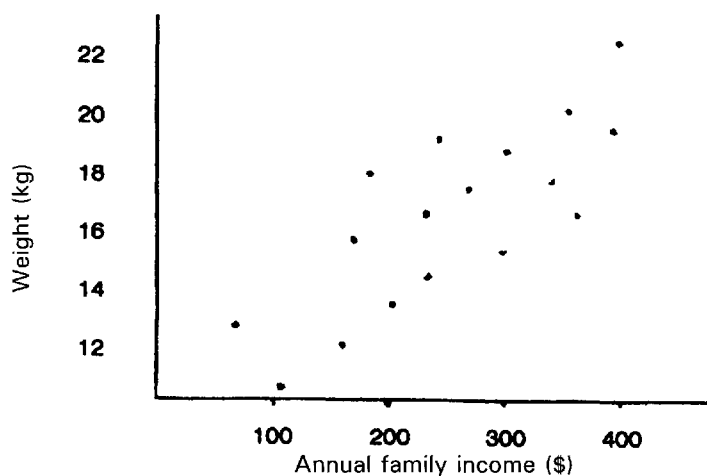


It is easy to show two or more distributions in one graph, as long as the difference between the lines is easy to distinguish. Thus it is possible to compare frequency distributions of different groups, i.e., the age distribution between males and females, or cases and controls.

5. Scatter diagrams

Scatter diagrams are useful for showing information on two variables which are possibly related. The example of a scatter diagram given below is used in **Module 31**, where we are dealing with the concepts of association and correlation.

Figure 22.5: Weight of five-year-olds according to annual family income



Note:

It is important that all figures presented in your research report have numbers, clear titles and clear labels (or keys).

In addition to the figures above, the use of **maps** may be considered to present information. For instance, the area where a study was carried out can be shown in a map. If the study explored the epidemiology of cholera, a map could be produced showing the geographical distribution of cholera cases, together with the distribution of protected water sources, thus illustrating that there is an association. If the study related to vaccination coverage, a map could be developed to indicate the clinic sites and the vaccination coverage among under-fives in each village, perhaps showing that home-clinic distance is an important factor associated with vaccination status.

V. MEASURES OF CENTRAL TENDENCY

Frequency distributions and histograms provide useful ways of looking at a set of observations of a variable. In many circumstances, it is essential to produce them to understand the patterns in the data. However, if one wants to further summarise a set of observations, it is often helpful to use a measure which can be expressed in a single number.

First of all, one would like to have a measure for the centre of the distribution. The three measures used for this purpose are the **MEAN**, the **MEDIAN** and the **MODE**.

1. Mean

The **MEAN** (or arithmetic mean) is also known as the **AVERAGE**. It is calculated by totalling the results of all the observations and dividing by the total number of observations. Note that the mean can only be calculated for numerical data.

Example 6:

Measurement of the heights of 7 girls gave the following results:

141, 141, 143, 144, 145, 146, 155 cm (a total of 1015 cm for 7 measurements)

The mean is thus $1015/7$, which is 145 cm.

2. Median

The **MEDIAN** is the value that divides a distribution into two equal halves.

The median is useful when some measurements are much bigger or much smaller than the rest. The mean of such data will be biased toward these extreme values. Thus the mean is not a good measure of the centre of the distribution in this case. The median is not influenced by extreme values. The median value, also called the central or halfway value, is obtained in the following way:

- List the observations in order of magnitude (from the lowest to the highest value or vice versa).
- Count the number of observations (n).
- The median value is the value belonging to observations number $(n + 1) / 2$ if n is odd or the average of the middle two numbers.

Example 8:

The weights of 7 pregnant women are 40, 41, 42, 43, 44, 47, 72 kg.

The median value is the value belonging to observation number $(7 + 1)/2$, which is the fourth one: 43 kg.

Note that the mean weight of this set of observations is 47 kg. This is an illustration of how the mean is affected by extreme values (in this case 72 kg) while the median is not. If the largest weight in this set of observations had been 51 kg instead of 72 kg, the median would still have been 43 kg, but the mean weight would have been 44 kg.

Note also that if there would be 8 observations: 40, 41, 42, 43, 44, 47, 49 and 72, the median would be 43.5 kg (the average of 43 and 44); the mean in this case would be 47.25 kg.

3. Mode

The MODE is the most frequently occurring value in a set of observations.

The mode is not very useful for numerical data that are continuous. It is most useful for numerical data that have been grouped.

In **Example 4** (number of patients treated for malaria at clinics) the mode is '0 to 19', as this outcome is recorded most frequently (25 times out of 80).

The mode can also be used for categorical data, whether they are nominal or ordinal.

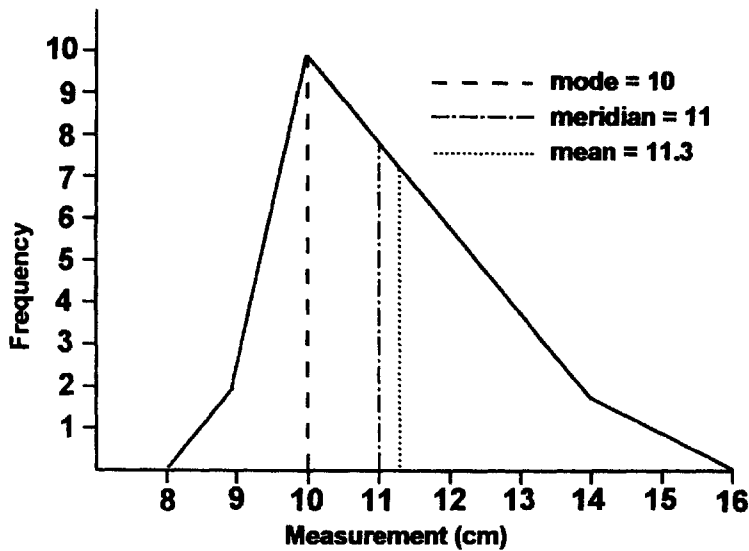
In **Example 1** (method of family planning) the mode is 'none'. In **Example 2** (number of clinics experiencing drug shortage) the mode is 'rarely'.

In summary, the mean, the median and the mode are all measures of central tendency. The mean is most widely used. It contains more information because the value of each observation is taken into account in its calculation.

However, the mean is strongly affected by values far from the centre of the distribution, while the median and the mode are not. The calculation of the mean forms the beginning of more complex statistical procedures to describe and analyse data.

Figure 22.6 shows a distribution curve in which the mean, the median and the mode have different values.

Figure 22.6: Mean, median and mode in a distribution curve.



GROUP WORK

1. Describe your sample(s) in terms of background variables (sex, age, etc). and dependent variables (e.g., defaulter/complier, user/non-user).
2. Make sure that you have made frequency counts for all variables in your study (from your data master sheets). Calculate percentages in relation to the total number of study units (or calculate proportions/ratios/rates where appropriate).
3. Check your objectives to determine which variables require frequency tables that should be included in your report. Usually frequency tables are presented for some of the background variables, the dependent variable(s) and the most important independent variables. Prepare the frequency tables.
4. Make histograms, bar charts, pie charts and/or line graphs, if useful. Prepare brief descriptions that interpret what each of the figures means.
5. Calculate means, medians and modes, if applicable, and interpret the results.
6. Familiarise yourself with the results and try to understand as fully as possible what they mean.

REFERENCES:

All epidemiology and statistics textbooks mentioned in **Modules 9** and **28**.

Module 22: DESCRIPTION OF VARIABLES

Timing and teaching methods

1 hour	Introduction and discussion
3 hours+	Group work

Introduction and discussion

- It is likely that the participants will be familiar with some of the concepts introduced in this module, such as percentages and proportions. Moreover, at this stage, the groups will have already prepared frequency distributions (including calculation of percentages). Therefore these concepts should be only briefly mentioned in the presentation, especially if the knowledge level of the participants is high, so as not to lose their interest. However, special attention should be given to what to do with missing values when calculating percentages.
- Although definitions of percentage, proportion, ratio and rate are given in the module, it is more important to provide examples or ask participants to provide their own.
- When presenting **Example 3**, you might also ask participants to describe how they have grouped numerical data and discuss whether there were too few or too many categories.
- Examples should not be merely presented; they should be used in informal exercises. For example, ask participants what is the mean, median and mode of a given set of measurements, instead of providing them with the answers.

Group work

- Before the group makes frequency counts for all variables from the master sheets, have them review whether the data have been categorised correctly. Also be sure that the total number of informants (study units) for each group that has been studied has been defined.
- Remind the group that fully developed frequency tables are only for those variables that have to be described in the final report. Usually tables are needed for some of the background variables of the target group(s), and sometimes for the most important independent variables. Many of the other background and independent variables will be presented using cross-tabulations (**Module 24**).

This page intentionally left blank

**Designing and Conducting Health Systems Research Projects
Part II: Data Analysis and Report Writing**

Module 23

ANALYSIS OF QUALITATIVE DATA

Steps in data analysis and report writing

Questions you must ask	Steps you will take*	Important elements of each step
What data have been collected for each research objective? Are data complete, accurate?	Prepare data for analysis	Review field experience Make an inventory of data for each objective/study population Sort data and check quality Check computer outputs (21)
What do the data look like? How can the data be summarised for easy analysis?	Summarise data and describe variables/identify new variables	Frequency tables, figures, means, proportions, descriptive cross-tabulations, (<i>quantitative data</i>) (22, 24); Coding, listing, summarising data in compilation sheets, matrices, flow charts, diagrams and narratives (<i>qualitative data</i>) (23)
How can the associations between variables be determined?	Analyse associations	Analytic cross-tables (24) Measures of association based on risk(25) Dealing with confounders (26)
	Prepare for statistical analysis	Measures of dispersion, Normal distribution and Sampling variation (27)
Do we measure differences or associations between variables?	Determine the types of statistical analysis	Choosing significance tests (28)
How can differences between groups be determined?	Analyse unpaired and paired observations	t-test, chi-square test (29) ** paired t-test, ** McNemar's chi-square test (30)
How can the associations between numeric variables be determined?	Implement measures of association	** Scatter diagram, ** Regression line and ** Correlation coefficient (31)
How should the report be written?	Write the report and formulate recommendations	Prepare outline for report Present and interpret data Draft and redraft Discuss and summarise conclusions Formulate recommendations (32)
How should the findings and recommendations be communicated, disseminated and used?	Present summaries and draft for implementation of recommendations	Discuss summaries and plan for implementation with all stakeholders (33)

* These steps need not be in the sequence in this diagram. **The sequence may be adjusted according to the needs of the research teams.**

** These elements are optional and may be omitted if not relevant for research teams

Module 23: ANALYSIS OF QUALITATIVE DATA

OBJECTIVES

At the end of this session you should be able to:

1. **Describe** efficient ways of ordering and summarising qualitative data.
2. **Indicate** why it is essential to start summarising and analysing during the field work.
3. **List** the major steps in analysing qualitative data and drawing conclusions.
4. **Make an outline** of how you will proceed with the ordering and summarising of your qualitative data, and with the subsequent analysis.
5. **Plan** on how to report your qualitative data, integrated in the most effective way with your other data.
6. **Indicate**, either now or at the end of data analysis, what additional activities you will undertake to test or confirm your findings in order to prove their validity.

I. Introduction

II. Procedures for processing and displaying qualitative data

III. Drawing and verifying conclusions, using different data sets

IV. Reporting the data

V. Further strategies for testing or confirming findings to prove validity (optional)

I. INTRODUCTION

In previous **Modules (9, 10, 13)** it was pointed out that we use qualitative research techniques if we wish to obtain insight into certain situations or problems concerning which we have little knowledge. Qualitative techniques such as the use of **loosely structured interviews with open-ended questions, (focus) group discussions, observations, projective and participatory approaches** will therefore be appropriate in many studies, especially at the onset. For sensitive topics they may be the only reliable techniques.

Irrespective of how and for what purpose the data has been collected, the researcher usually ends up with a substantial number of pages of **written text** that needs to be analysed.

Although procedures and outcomes of qualitative data analysis differ from those of quantitative data analysis, the principles are not so different. In both cases the researcher will have to:

- describe the sample populations;
- order and reduce/code the data (data processing);
- display summaries of data in such a way that interpretation becomes easy, e.g., by preparing compilation sheets, flowcharts, diagrams or matrices;
- draw conclusions, relate these to the other data sets of the study and decide how to integrate the data in the report; and
- if required, develop strategies for further testing or confirming the (qualitative) data in order to prove their validity.

We will now examine each of these points in more detail.

II. PROCEDURES FOR PROCESSING AND DISPLAYING OF QUALITATIVE DATA

1. Description of the sample population in relation to sampling procedures

A useful first step in data processing (as well as in the reporting of findings) is a description of the informants. If numbers allow, relevant background data may be tabulated, for example on age, sex, occupation, education or marital status, as is the practice in quantitative studies.

However, as qualitative data originates from small samples (sometimes a handful of key informants or focus group discussions and observations) more information is required to place the data in its context.

For example, who were the key informants, what made you decide to choose them? Who took part in the focus group discussions? How were the participants of the groups selected and how representative are they for your study population? For observations: under what circumstances were they carried out? Who were observed, and by whom?

Unless this type of information is provided, interpretation of data may appear haphazard.

2. Ordering and coding of data

We will discuss two types of qualitative data:

- answers to open questions, and
- more elaborate narratives from loosely structured interviews or FGDs.

(1) Answers to open questions

The most commonly collected qualitative data are the answers to open questions. They form part of every HSR study. When developing your protocol, in **Module 13**, you already did an exercise in the systematic ordering of such data: on the answers to the question 'Why are you smoking?' which we will discuss in depth again to analyse the different steps*.

- (1) A first, basic step in the analysis of answers to open questions is to **list** the answers of a *sample of 20-25 informants* as they were provided (adding the questionnaire number in order to avoid losing the connection with the informant's other data).
- (2) Then **read** the answers carefully, remembering the purpose of the question. The question 'why are you smoking' was supposed to help nursing students to develop an intervention against smoking.
- (3) **Make rough categories** of answers that seem to belong together and **code** them with a key word. For example, answer 3 (It gives me pleasure) and answer 14 (I like to blow smoke rings) could be labelled with the term 'pleasure', which could be abbreviated with the code *pleas*.
- (4) Then **list** again all answers but now **per code**, so that you get some 5-7 short lists, for example:

<i>Pleasure (pleas)</i>	<i>Being sociable (soc)</i>	<i>Giving self-confidence (selfc)</i>
2. I like the feel of the cigarette in my hand	10. All my friends are smokers	6. Because I feel confident and in-charge when smoking
3. Because it gives me pleasure	11. It helps to make people more friendly and comfortable, when offering a cigarette	7. It helps me to think better
5. I like to blow the smoke through my mouth and nose		18. It helps me to reduce the pressure and tension at work
14. I like to blow smoke rings		17. It helps me to relax
15. I like the taste		

- (5) Then **interpret** each list, and end up with some 5-7 meaningful categories with a characteristic key word. For example: *Pleasure, being sociable, giving status, giving self-confidence, addiction, defiance*. There may be discussion on the need to split up some categories or combine others with few answers. Answers 17 and 18, for example could be put in a separate category *reducing stress*. In that case there would be seven categories. The category *defiance* may have two answers: 4. I do not see why I would give up smoking!! and 12. Why not?!! The exclamation marks indicate that defiance rather than lack of knowledge forms the motivation for the answer. Without this addition by the interviewer, these answers would have been difficult to code.

Now you can make a tentative interpretation according to the assumed willingness of your informants to change their behaviour. For those who smoke for *pleasure* or to *socialise* it might be most easy to give up smoking. Those who are *addicted* but tried to stop and those who feel they derive *status* from smoking might form a middle category, whereas for those who smoke to enhance their *self-confidence* and reduce *stress* or who are very *defiant* at the question why they smoke, it might be most difficult to stop.

* These steps have been adapted from Willms and Johnson (1996).

- (6) Now try a next batch of 20-25 answers and **check if the labels work**. It is well possible that at this stage still some labels will be changed or that you decide to add new categories or combine others.
- (7) **Make a final list** of labelled categories and code all data including the data you already processed with the abbreviated codes.

Then discuss whether you will stick to your tentative interpretation of the data and what this means for the content of the messages to address different reasons for smoking. This *content analysis* is the most important purpose of the analysis. By *counting* the answers under each label, however, the researcher will gain insight as well in how *common* the different reasons are.

(2) Elaborate narratives

The data from interviews with key informants or focus group discussions (FGDs) are as a rule more bulky than answers to open questions. The carefully transcribed field notes and tapes (see **Module 10.C** on FGD and **Module 13**) may consist of pages of narrative text. When analysing the texts we usually discover that, no matter how good our guidelines for the discussion were, the data contain valuable information but also a number of less essential details. In addition, the data is usually not presented in the order we need for our analysis, since informants may jump from one topic to the other.

To make the analysis easier, we have to **order** and **reduce** the data. Ordering is best done in relation to the objectives and the discussion topics. Again, it is best to systematically follow a number of steps.

- (1) **Reread** your objectives and discussion topics
- (2) Carefully **read** a number of the interviews, FGDs or narrative observations you want to process. Number the material according to the broad discussion topic it pertains to. Use a yellow marker to highlight particularly illustrative remarks. Use the margins to define sub-topics.

For example, in a gender and leprosy study carried out in different countries (used as example in **Modules 4, 8** and **11**) it appeared that the discussion topic *stigma* had to be differentiated according to different social settings in which it occurred: among close relatives (parents-children), spouses, in-laws, and community members. Further, a distinction had to be made between self-stigmatisation (e.g., a wife diagnosed as a leprosy patient encouraging her husband to marry a second wife in order to prevent divorce, or a patient not attending community meetings for fear of being avoided) and stigmatisation by others. Different degrees of severity in stigmatisation could also be distinguished, varying from slight avoidance to complete expulsion. If stigma would be topic (11) in your discussion list, you would mark everything related to stigma with an (11) in the margin, and add *key words* such as *self-stigm.*, *spouse*, *in-laws*, *comm.*, in the margin, as well as key words such as *sleep(ing) sep(arately)* or *divorce* indicating the severity of the stigma. (See **Annex 10C.2** in Module 10C for an example.)

- (3) **List all key words that belong to a certain topic in the sub-categories** that have been developed under (2). E.g., everything belonging to stigma could be subdivided and listed in the four major social settings in which stigma was found to manifest itself.
- (4) **Interpret** the data, e.g., distinguish the major forms in which stigma manifests itself in these different social settings, try to make a ranking order of severity and link it to other variables (such as degree of deformity, socio-economic status) in order to understand differences in stigma.
- (5) Then **code all your qualitative data** in this way. If necessary, adapt your coding scheme as you order, code and interpret more data. In that case, you should again read and possibly re-code the material you have already processed.

Note:

You may already have analysed and coded your qualitative data in the field (as advised in **Module 13**), in order to adjust and deepen your interview guides or topic lists. In that case it may be possible to develop your final coding list in one cycle instead of two.

However, instead of developing a very detailed coding system on your rough data, you may also refine your interpretation as you record your roughly coded, summarised data in **COMPILATION SHEETS**.

3. Summarising data in compilation sheets

After ordering the data we will have to summarise them. A useful first step is summarising all data of each study unit per study population on separate compilation sheets.

Like the master sheets for quantitative data, compilation sheets for qualitative data consist of a number of columns with the topics covered by the study as headings. These may be further sub-divided in smaller themes that you identified and coded when ordering the data (see **Annex 23.1**). Each interview, FGD or observation gets a number and is successively entered in that sequence on the relevant compilation sheet. If there are different categories of informants within one study population, for example, young mothers and an older generation of mothers, or male and female patients, the data for these groups are entered on separate sheets. If the topics covered in those sub-groups are not completely identical, it is important to be systematic and follow roughly the same sequence of topics for each category of informants. The information inserted is summarised in key words and key sentences, clear enough to remember the statements informants made. (As the number of each study unit is entered in the compilation sheet, it is always possible to go back to the original data and present the full statement, for example in a presentation or in the research report).

Now you have an overview of all data per study population on one or more big sheet(s). If you read the columns, you have a list of answers of all group members on a certain (sub-)topic. If you read horizontally, you can per informant relate different topics to each other or to personal characteristics of the informant. It becomes also easy to compare the answers of different groups on specific issues by comparing compilation sheets.

For example, in **Annex 23.1**, the personal data of leprosy patients (recently declared cured) and a number of topics and sub-topics discussed with them are presented. Stigma actually experienced, which originally was one topic, has in the compilation sheet been subdivided in the four major social settings in which stigmatisation may occur: close blood relatives, marriage, wider circle of spouse's relatives and community. In each of those still finer distinctions can be made (e.g., community can be neighbours, friends, work mates, school mates or distant community members). As samples are small, these may all be inserted under the heading 'community'. Codes (*italics*) can be added to the statements presented in key words, for example *big fear* and *worried* under the heading 'first reaction'. From the three examples presented, it already appears (confirmed by the analysis of all data in all four countries) that in general the stigma feared when patients hear the diagnosis of leprosy is bigger than the stigma in reality experienced. Patient (12) is in this respect an exception. Ironically, the husband who divorced her had already died from another disease at the moment she was declared cured from leprosy. Horizontal comparison of the data of patient (1) teaches us that it is highly unlikely that the man's friends do not know about the disease, as even after he has been declared cured he has visible signs. Here the researchers had to interview the friends to find out if indeed this man was (or had not been) stigmatised at all by the community.

You may notice that interpretation of data and labelling becomes indeed easy when using compilation sheets, as a researcher can visualise all aspects of his/her informants even if (s)he looks at one aspect at a time for the whole study population.

A next step in summarising may be the combination, contrasting or further analysis of important topics through graphical displays such as matrices, diagrams, flow charts and tables.

4. Further summarising of data in matrices, figures and tables

Matrices

Matrices can be used for quantitative as well as qualitative data comparison. In qualitative data we may compare different groups or data sets on important variables, presented in key words.

A MATRIX is a chart that looks like a cross-table, but contains words (as well as, sometimes, numbers).

In a focus group discussion on changing weaning practices, the researchers listed the answers of young mothers concerning the introduction of soft foods and those of mothers above childbearing age. They then summarised these answers in a matrix:

Figure 23.1: Matrix on introduction of soft baby foods among mothers of different age groups

AGE GROUPS	ONSET SOFT FOOD	TYPE OF FOOD	FREQUENCY OF SF/DAY
Young mothers (20-30 years)	Range: 4-7 months Average: 6 months	<ul style="list-style-type: none"> • Soft porridge • Soft porridge with pounded groundnuts • Mashed potatoes, mashed fruits, soaked biscuits 	2-4 times daily <ul style="list-style-type: none"> • Depends on availability of mother and caretaker • Depends on appetite of child
Mothers past child-bearing age (> 45)	Range: 5-11 months Average: 8.5 months	<ul style="list-style-type: none"> • Soft porridge • Soft fruit 	1-2 times daily <ul style="list-style-type: none"> • Depends on availability of mother and caretaker • Depends on appetite of child

This type of display made it easy for the researchers to conclude that:

- younger mothers start giving soft foods, on average, 2.5 months earlier than the generation of their own mothers;
- younger mothers use a larger variety of soft weaning foods than women in the preceding generations; and
- younger mothers give soft foods to their babies more frequently, but for the same reasons as their mothers did.

Matrices facilitate data analysis considerably. They are the most common form of graphic display of qualitative data. They can be used to order and compare information in many ways, for example, according to:

- time sequence (of procedures being investigated in different periods, for example),
- type of informants (as in the example above), or
- location of data collection (to visualise differences between rural and urban populations).

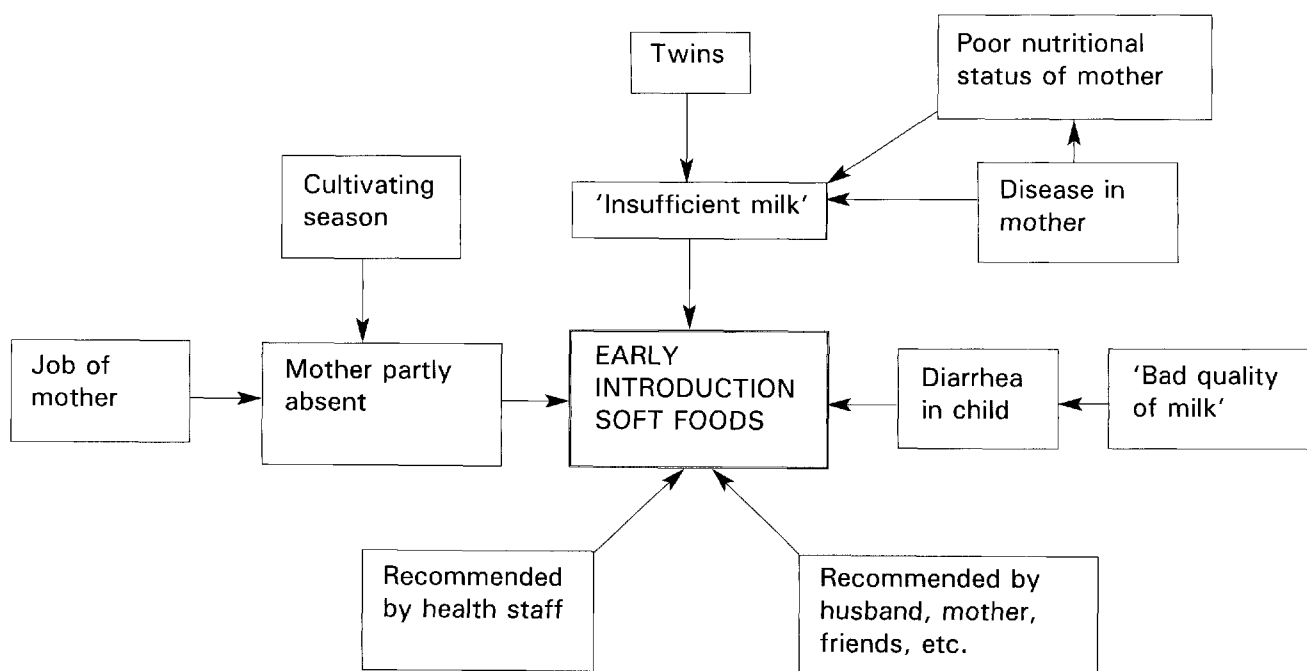
Diagrams

A **DIAGRAM** is a figure with boxes containing variables and arrows indicating the relationships between these variables.

When analysing the problems you wanted to investigate during the development of your protocols, most groups developed a diagram. In a similar way diagrams can be developed to summarise findings of a study. (See **Figures 23.2** and **23.3**).

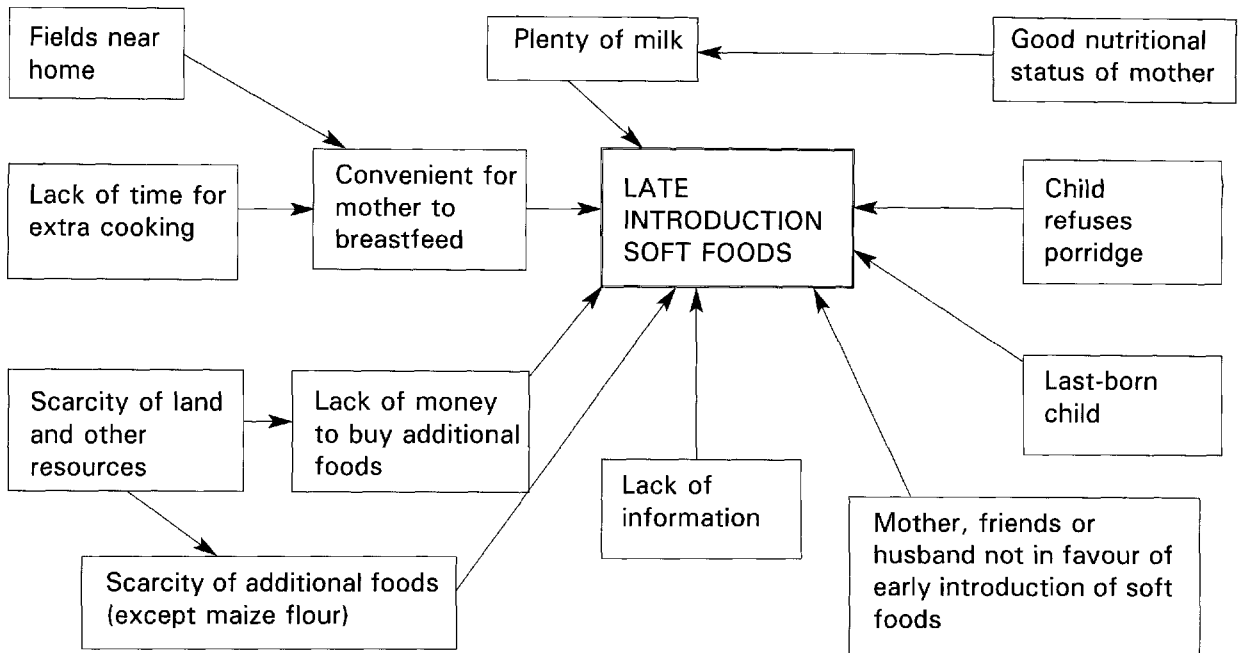
You might use a diagram to illustrate a crucial issue in your study, combining all available qualitative and quantitative data collected.

Figure 23.2: Reasons for early introduction of soft foods by young mothers



Diagrams, like matrices, can be of great assistance in providing an overview of the data collected and in guiding data analysis.

Figure 23.3: Reasons for late introduction of soft foods by young mothers



Flow charts

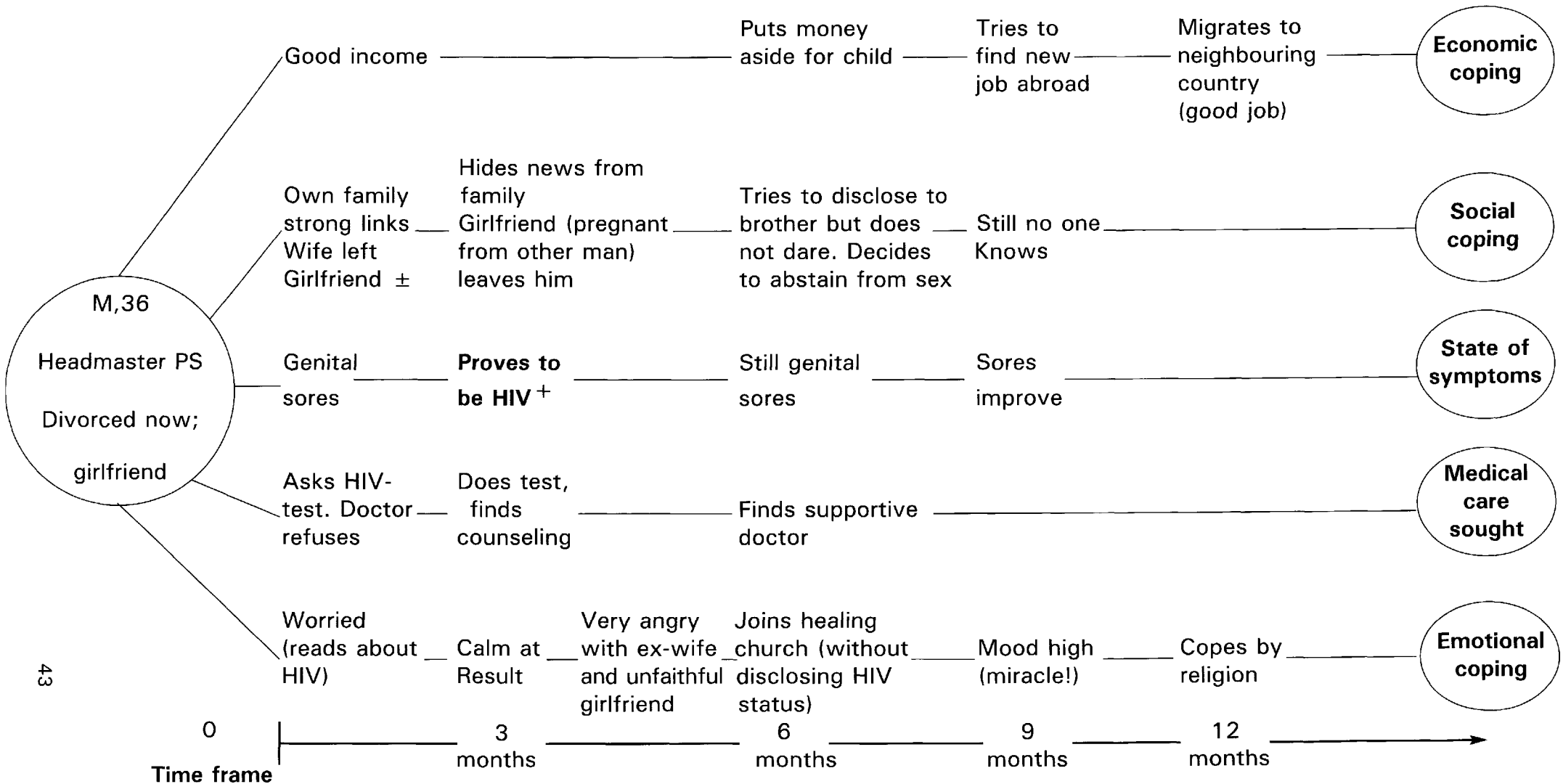
FLOW CHARTS are special types of diagrams that express the logical sequence of actions or decisions.

The figure preceding **Modules 1-18**, indicating the successive steps in protocol development, is an example of a flow chart.

Flow charts are especially useful to summarise different flows of events that are mutually connected. A counselling team in Bulawayo, Zimbabwe, for example, which interviewed some 95 HIV positive persons in-depth over a period of two years, summarised the roughly 100 pages of interview material for each informant by drawing five lines (see **Figure 23.4**). One central line presented the development of the disease over time, with crises and periods of relative well-being. Another line presented different forms of medical care sought, a third the flaws in economic status connected to the disease (e.g., loss of job, seeking employment elsewhere), a fourth the possible changes in social status such as divorce or (re)marriage, whereas a fifth line presented the patient's emotional status linked to events occurring in the four other fields (e.g., positive coping, depression). These flow charts were extremely useful for comparison of data, per informant and between different groups of informants (e.g. males/females, single/married). They highlighted the impact of the disease on the lives of different groups of patients and their way of coping with it.*

* Meursing (1997) *A world of silence*.

Figure 23.4: Flowchart on coping of HIV+ persons with their condition over time



43

Example of relatively well-to-do man who copes in solitude despite supportive relatives because he is too ashamed to unclose his HIV+ status.
Adapted from Meursing K (1997) *A world of silence; Living with HIV in Matabeleland, Zimbabwe*. Amsterdam: Royal Tropical Institute.

Tables

A TABLE is a chart with rows and columns that has numbers in the various cells or boxes.

Qualitative data can also be categorised, coded, inserted in master sheets or computer and *counted*, together with other quantitative data, and displayed in tables. Answers to open-ended questions in questionnaires will usually be categorised and summarised in this way. However, you will in the first place want to *analyse the content* of the individual answers in each category. (See section II-2 and section III in this module.)

III. DRAWING AND VERIFYING CONCLUSIONS

Drawing and verifying conclusions is the essence of data analysis. It is not an isolated activity, however. When we start summarising our data in compilation sheets, flowcharts, matrices or diagrams, we continuously draw conclusions, and modify or reject quite a number of them as we proceed. Writing helps generate new ideas as well. Therefore **writing should start as early as possible**, right from the onset of data processing and analysis, if only for ourselves. No creative insights should get lost!

Note:

Collection, processing, analysis and reporting of qualitative data are closely intertwined, and not (as is the case with quantitative data) distinct successive steps. It may often be necessary to go back to the original field notes and verify conclusions, collect additional data if available data appear controversial, and get feedback from all parties concerned.

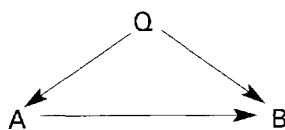
Identifying variables and associations between variables

In **Module 8** we stated that sometimes we do not know enough about a situation to define variables beforehand. Only during or at the end of the study it will be possible to define certain variables and search for associations with other variables, without having the prior aim of *measuring* them. Many HSR studies have qualitative parts with open questions, key informant interviews, focus group discussions or observations for the purpose of identifying these variables. The researcher who uses such a qualitative approach should be like a detective who searches for evidence, accounts for countervailing evidence, and verifies the findings by looking for independent, supporting evidence, until (s)he is confident about possible associations among certain variables which shed light on the problem under investigation.

For example, if we find among the mothers who wean their children early that quite a number have jobs, we may assume that having a job contributes to early weaning. Similar studies carried out elsewhere with similar findings support this assumption (independent evidence). Only if there are very few employed women who wean their children late, however, can we be more certain that our assumption is true, and for each of those exceptions we should try to find an explanation. Do the mothers take their children with them (crèche at place of work) or do they work near their homes so that they can feed the baby during breaks? Or do they successfully combine breast-milk with alternatives? If yes, why don't more mothers try this combination? etc. etc.

Finding confounding or intervening variables

Sometimes variables appear to be related but the association cannot easily be explained. Other times it seems that variables should logically go together, but you cannot find a relationship. In cases such as these there may be another variable ('Q') influencing the association between the two variables concerned, that has to be identified (see **Modules 8, 9 and 26**).



For example, one expects a relationship between the quality of drinking water and the incidence of diarrhoea. It is assumed that the incidence of diarrhoea would decrease as the number of water faucets in a village increased. If there is no change over time, there might be a confounding variable. People, for example, may dislike the taste of tap-water so much that they use it for everything, except for drinking.

Note:

Such unexplained associations may appear in any study. The essential characteristic of a qualitative research approach is that it purposively looks for such associations **during** the fieldwork, and that additional questions and tools may be developed to highlight such relationships. In quantitative surveys that attempt to objectively measure the strength of a *presupposed* association between two variables, the tools should not be changed once the fieldwork is ongoing.

Integrating qualitative and quantitative data

Thus far we have discussed the analysis of qualitative data as a separate activity. However, if a research team has collected qualitative as well as quantitative data, which is the case in most HSR studies, it would be foolish not to look at them in combination, as this can inspire to deeper and more rewarding analysis.

For example, the Indonesian 'gender and leprosy' research team found, when analysing the registration data of 4500 new leprosy patients who had registered over the past five years, that the M/F ratio was most unfavourable in the age group of 15-44 years. This was a puzzling finding, as in Nepal women in this age group were reporting much better (though still less than men). In-depth interviews with staff revealed that they suspected adolescent girls and young women to hide their skin patches, because of shameful associations with dirt, ugliness. This provided the incentive for a further break down of the quantitative data, which revealed that the M/F difference in reporting was indeed most pronounced in the 15-34 age group, and levelled off above 35. The reason(s) for this relatively large gender difference in the younger age groups were then further explored.

Content analysis of qualitative data for action

Quantitative data serve in the first place to convince health authorities **that** there is indeed a serious, sizeable problem; qualitative data help to provide ideas on **how** to solve it. The FGDs on weaning foods with young mothers and mothers who had surpassed the childbearing age, for example, will yield many suggestions on how to develop interventions with the mothers which they are likely to consider useful and will be able to implement. Likewise, the in-depth interviews with leprosy and ex-leprosy patients will provide new insights into how best to counsel new patients and their close relatives/spouses in order to reduce unnecessary fears.

Computer analysis of qualitative data

With the ever-increasing importance of computers in research, strategies for analysing qualitative data by computer have been/are being developed. There are several possibilities, ranging from simple word processing programs to highly sophisticated Qualitative Data Management Software including possibilities for statistical testing of associations. As numbers are usually small in HSR and content analysis, which can be done by hand, is most likely more important than *testing* of associations, we will not elaborate these techniques here. Rather we refer the interested students to Anthropology or Psychology Departments at universities that have experience with programs such as Qualitan or SPSS for qualitative data processing.

IV. REPORTING THE DATA

Basically, there are two ways of reporting qualitative data that form part of a study in which different research techniques were used. One way is summarising the major qualitative results in a separate section of the findings, with examples and quotations, following the objectives that guided the collection of this particular data. The results would then be discussed in the chapter 'Discussion', together with the results of other, more quantitative data collection tools and would subsequently be reflected in the summary of the findings and the recommendations.

Another possibility is to fully integrate different data sets in the chapter of findings, ordered according to the objectives of the entire study. If quantitative and qualitative data have been analysed and sometimes even collected in an integrated way, it would also be logical to present them in an integrated fashion. Attention should be paid that no valuable data get lost. Therefore a rough draft of all important findings is required in any case, after which can be decided to present the data either in separate sections or chopped up for integration with other data. (For details see **Module 32**.)

V. FURTHER STRATEGIES FOR TESTING OR CONFIRMING QUALITATIVE FINDINGS TO PROVE VALIDITY

Researchers who use quantitative research designs reduce their data to numbers and apply statistical tests. This does not necessarily insure that their research results are valid: something may have gone wrong during sampling or collection of data or even in the earlier design of the study (overlooking possible confounding variables). The following strategies will therefore be of use to any researcher. They are particularly relevant, however, to qualitative research, since the small numbers of qualitative data often generate questions concerning its validity.

1. Check for representativeness of data.

Although in qualitative research informants have usually not been selected randomly, they must have been selected systematically, according to previously established rules. (See **Module 11**.) Check whether you have indeed interviewed all categories of informants needed to get a complete picture of your topic (not relying excessively on talkative authorities). Make sure that you do not generalise from unrepresentative events.

2. Check for bias due to observer bias or the influence of the researcher on the research situation. We discussed this in detail in **Module 10**.

3. Cross-check data with evidence from other, independent sources.

These sources may be different independent informants, different research techniques employed to investigate the same topic, or results from other, similar studies. (See **Modules 5, 9 and 10**.) The data should confirm or at least not contradict each other.

Actively cross-checking data, looking for independent evidence or counter-evidence, is one of the most important ways to enhance the validity of research data.

For example, answers of husbands and wives (and other informants concerned) should confirm each other on such issues as who decides whether and what family planning methods should be used, who decides whether daughters should be circumcised, or what has changed in husband-wife relationships after the diagnosis of leprosy or another feared disease in one of the spouses.

4. Compare and contrast data.

Comparison will often have been built into the research design through including different categories of informants.

If we want to be sure, **for example**, that variable A (high level of education) influences variable B (use of family planning methods) we have to compare a group of mothers with high education to a group of mothers with low education on their use of family planning methods.

Comparing and contrasting data is important if you are attempting to **identify** your variables as well as to **confirm** associations among variables.

5. Use extreme (groups of) informants to the maximum.

In the discussion of study design and sampling we stated that it may be useful to look for categories of informants that represent the extremes on a certain variable.

For example, you may find it most useful to study 'drop-outs' and regular attendees of TB services, leaving out the category of irregular attendees. This may be the most efficient way of identifying the key variables that influence the attendance behaviour of TB patients.

6. Do additional research to test the findings of your study.

The results of your study may be so intriguing that you decide to do a follow-up study afterwards. Such a study may be undertaken for several reasons:

- to replicate certain findings,
- to rule out (or identify) possible intervening variables,
- to rule out rival explanations by investigating them, or
- to look for negative evidence.

Additional studies undertaken for one or more of these reasons may serve to make the results of your original study more convincing.

7. Get feedback from your informants.

Throughout **Modules 1 - 20** we have stressed that you need to involve all parties concerned in the various stages of the research. This is important not only for ethical reasons or because it will improve the chances that the results will be implemented, but also because **it will improve the quality of your study design, of your data, and of the conclusions drawn from these data**. Suggestions and additional information collected during feedback sessions will invariably increase the quality of your research report.

GROUP WORK (Time needed will depend on the amount of qualitative data collected.)

1. **Check whether you listed all sources of qualitative data** for each objective when, in the group work session of **Module 21**, you made an inventory of all your data.
2. **Ensure** you have finished the **categorisation of answers to all your open questions**, included them in your master sheets or computer analysis together with other data sets, and that you then do a content analyses of the answers for inclusion of relevant data in recommendations for action or subsequent action plans (see **Module 33**)
3. If your study included FGDs, interviews with key informants or observations: **describe your samples**.
4. Organise this data by topic, further code it, if necessary, **and enter the data by topic on compilation sheets**.
5. **Decide** whether you will use matrices, diagrams and/or flowcharts to summarise your data.
6. **Interpret the data**, comparing different groups of FGD participants or key informants (if you have them) and see how they answer your research objectives.
7. **List the major findings and conclusions of the qualitative data** and determine how they complement data from other sources in your study.
8. **Decide on how you want to enter the data in your report**: either in one section or integrated with the findings collected through other data-collection techniques. Decide what should come in the discussion and what is useful material for developing interventions.
9. **Verify your conclusions** (see section V) and decide whether and how you would like to further test certain conclusions.

REFERENCES:

- Miles MB and Huberman AM (1984) *Qualitative data analysis, a sourcebook of new methods*. Beverley Hills, CA, USA.: Sage Publications.
- Patton MQ (1990) *Qualitative Evaluation and Research Methods*. 2nd ed. Newbury Park, CA: Sage Publications.
- Spradly JP (1979) *The ethnographic interview*. New York, NY, USA.: Holt, Rinehart and Winston.
- Walker R (ed) (1985) *Applied qualitative research*. Hants, UK: Gower Publishing Company Ltd.
- Willms DG and Johnson NA (1996) *Essentials in Qualitative Research: A Notebook for the Field*. Hamilton, Canada: Mc Master University.
- Yin RK (1984) *Case study research: design and methods*. Beverly Hills, CA, USA.: Sage Publications.

NB: A major source of inspiration for writing this module was Miles and Huberman's book. Section V of this module is a heavily abbreviated and adapted version of their chapter VII.

Annex 23.1: Example of compilation sheet (gender and leprosy)

Nr	Personal data					Symptoms		First reaction(10)	Stigma in reality experienced (11)				Ec./domestic act (12)	Perception of cure (25)
	Sex	Age	Educ.	Marr.	Ec. status.	at diagn. (6)	now (18)		Spouse	Relatives	In-laws	Comm.		
1	M	40	6yrs	Yes	Farmer Shopkeeper	<ul style="list-style-type: none"> • Patches • Painful nerves • Dropfoot (2) 	<p>↓ None</p> <p>Still</p>	<p><i>Big fear</i></p> <ul style="list-style-type: none"> • Wife will run away • Community will isolate him • Fingers and toes will drop off • No longer able to work and sustain family 	<ul style="list-style-type: none"> • Remains <i>supportive</i> • Helps more in shop • He decided to abstain from sex (8 months) <i>Self-stigm.</i> 	<ul style="list-style-type: none"> • Children <i>supportive</i>; small ones not aware • Parents & Br/Si visit + mix as before 	Not told, <i>hiding</i>	<ul style="list-style-type: none"> • Not told • Thinks friends don't know • Behave as before <i>Hiding. No stigma?</i> 	Hires labour (No force to farm) <i>Income</i> ↓	<i>No</i> (still signs: drop-foot)
10	F	21	8yrs	Yes	Hu. farms Fa. big farmer	Patches (teacher saw and referred her to HC)	<i>None</i>	<p><i>Knew little; worried</i></p> <ul style="list-style-type: none"> • Bad disease • Fiancé will break off marriage proceedings 	<ul style="list-style-type: none"> • Fiancé inquired at HC: - if curable - if she could get children • Marriage postponed till patches subsided • She now has child <i>Stigma reversed</i> 	<ul style="list-style-type: none"> • Parents very <i>supportive</i> 	First wife of hu. told in whole village	All villagers came to marriage. <i>No stigma</i>	Does everything	Yes (No signs)
12	F	60	-	Yes, but now divorced + trade	Small farming	Patches	<p>↓</p>	<p><i>Worried</i></p> <ul style="list-style-type: none"> • Bad disease • Hu. angry 	Hu. kicked her out. <i>Divorce</i>	Son took her in. <i>Supportive</i>	Da.-in-law <i>supportive</i>	Avoided big meetings but now OK <i>Self-stigma reversed</i>	Small trade to earn bus fee for treatment	Yes, cured (but hu. died!)

Module 23: ANALYSIS OF QUALITATIVE DATA

Time and teaching methods

1 hour	Introduction and discussion
1 hour	Group work

(Time should be adjusted, depending on amount and type of qualitative data.)

Introduction and discussion

- If none of the groups have qualitative data other than some open-ended questions in questionnaires, you may wish to concentrate on sections II, to give participants an overview of how one could process qualitative data, and only briefly touch on sections III and IV.
- However, if the course participants have experience/training/interest in research, you might fully cover sections III and IV, even if they have not collected large quantities of qualitative data. The procedures presented for drawing conclusions and testing validity are pertinent to all types of research, and the methods for checking and cross-checking data may not be known to all participants.
- Refer to the analysis diagram each group made when preparing its research proposal, to the flow chart in front of **Modules 1-18** and **22-33**, and present any other examples of charts or graphs you find illustrative on overhead sheets or flip charts.
- If one or more groups have done extensive qualitative research, cover the module in detail, using examples from their studies. Most likely none of the participants will be very familiar with analysis of qualitative data.
- Let the groups that have done qualitative research describe in plenary how they analysed the data from focus group discussions, observations, and/or interviews with key informants. Ask what additional questions they added or questions they dropped in the course of successive interviews, and why.

Group work

- **For all groups:**

Check whether any of their open-ended questions require analysis of the content of individual answers. Some opinion questions might provide valuable illustrative material for their reports. Take note of this, as the groups might forget such data when they get involved in tables and statistics. Discuss whether merely listing the statements is sufficient for content analysis or whether graphic display of the data would be desirable.

For those groups that have elaborate qualitative data from FGD or key informant interviews:
Review all the data available with them and assist them in following the group work directions.